THE ROYAL SOCIETY

Received 24 February 2004
Accepted 26 July 2004
Published online 23 November 2004

The impact of cross-immunity, mutation and stochastic extinction on pathogen diversity

Laith J. Abu-Raddad* **and Neil M. Ferguson**

Department of Infectious Disease Epidemiology, Imperial College London, St Marys Campus, Norfolk Place, London W2 1PG, UK

We examine the dynamics of antigenically diverse infectious agents using a mathematical model describing the transmission dynamics of arbitrary numbers of pathogen strains, interacting via cross-immunity, and in the presence of mutations generating new strains and stochastic extinctions of existing ones. Equilibrium dynamics fall into three classes depending on cross-immunity, transmissibility and host population size: systems where global extinction is likely, stable single-strain persistence, and multiple-strain persistence with stable diversity. Where multi-strain dynamics are stable, a diversity threshold region separates a low-prevalence, low-diversity region of parameter space from a high-diversity, high-prevalence region. The location of the threshold region is determined by the reproduction number of the pathogen and the intensity of cross-immunity, with the sharpness of the transition being determined by the manner in which immunity accrues with repeated infections. Host population size and cross-immunity are found to be the most decisive factors in determining pathogen diversity. While the model framework developed is simplified, we show that it can capture essential aspects of the complex evolutionary dynamics of pathogens such as influenza.

Keywords: infectious disease; mathematical model; antigenic variation; cross-immunity; population dynamics

1. INTRODUCTION

One of the key current research priorities for infectious disease epidemiology is understanding the dynamics of antigenically variable pathogens such as influenza, malaria, dengue, and meningitis (Cooper 2001; Grenfell & Gog 2001; Gupta & Maiden 2001; Grenfell *et al*. 2004). These diseases continue to cause millions of deaths each year and thus represent major public health challenges. Annual influenza epidemics cause upwards of 20 000 deaths in the USA alone (Simonsen *et al.* 1998), while malaria remains the single biggest killer of children under 5 years old (WHO/UNICEF 2003). Quantitative understanding of the epidemiology of such diseases—and thus of the likely impact of potential control measures—requires their multi-strain population structure to be taken into account (Gupta *et al.* 1994).

The critical feature of most multi-strain pathogens is that infection by one strain induces partial immunity to future infections by other strains. Ecologically, such cross-immunity acts as a competitive interaction between strains, which when coupled with the nonlinearity of underlying transmission dynamics leads to intense frequency-dependent selection within pathogen populations (Gupta *et al.* 1998).

However, until now, most theoretical studies of multi-strain systems have assumed a certain possible maximal number of strains, and examined how cross-immunity affects transmission dynamics or reduces diversity relative to the assumed maximal possible level, rather than examining the dynamics of systems with potentially arbitrary numbers of strains.

By contrast, here we analyse how cross-immunity, mutation and demographic stochasticity (caused by the finite size of host populations) constrain the observed diversity of pathogens with conceivably unlimited maximal diversity. Directly modelling the stochastic dynamics of host–pathogen systems with arbitrary pathogen diversity remains computationally impractical except in special cases. We therefore develop a new approach to examining the stochastic persistence of multi-strain systems based on gaining an analytical understanding of the equilibria of the symmetric deterministic multi-strain system, and a novel application of previous analyses of stochastic epidemic models (Nasell 1999).

A range of past work (Gupta *et al.* 1996, 1998; Andreasen *et al.* 1997; Lin *et al.* 1999; Gomes & Medley 2002; Ferguson *et al.* 2003) has examined the dynamics of multi-strain systems with cross-immunity using deterministic models. However, most of these analyses have focused on systems with relatively few strains because of the mathematical intractability of models of systems with many strains. This intractability arises from the need to track all possible infection histories to calculate the expected cross-immunity experienced by an individual exposed to a new strain, giving exponential growth in the number of state variables as a function of the number of strains modelled.

The mechanisms and action of cross-immunity are still not well understood for many host–pathogen systems, and may vary between pathogens. This uncertainty has led to a variety of modelling approaches being adopted (Ackerman *et al.* 1990). In some cases, cross-immunity has been assumed to reduce the infectiousness of future infections (Gog & Grenfell 2002; Gupta *et al.* 1996, 1998), while in others it reduces the susceptibility to future infections (Castillo-Chavez *et al.* 1989; Andreasen *et al.* 1996;

* Author for correspondence (l.aburaddad@imperial.ac.uk).

Lin *et al.* 1999). The subtle difference between the two is that for the former, individuals exposed to a strain always gain homologous immunity to it, while for the latter, no additional immunity is induced if exposure fails to result in infection. However, it appears that system dynamics are insensitive to which of the two formulations is used (Dawes & Gog 2002; Ferguson & Andreasen 2002; Gomes & Medley 2002).

Model simplification is possible (Gog & Grenfell 2002; Gog & Swinton 2002) if one assumes that cross-immunity acts on infectiousness, and in addition that cross-immunity is polarized, such that a proportion of the exposed population gain total immunity against heterologous infection, with the rest gaining none, rather than all the exposed population getting partial immunity. However, the extent of homology in system dynamics between models using the more realistic (though less tractable) assumption of partial cross-immunity and those using polarized immunity remain unclear; hence here we use the former and later comment on the comparison with the polarized immunity formulations.

2. CHARACTERIZING THE SYMMETRIC EQUILIBRIUM OF MULTI-STRAIN SYSTEMS

We use the n-strain susceptible–infected–recovered (SIR) model formulation derived by Andreasen *et al.* (1997). The model structure is complex, as we must allow for compartments that are recovered and immune from previous infections but simultaneously still susceptible to infections by other strains, where the level of susceptibility is determined by the cross-immunity profile. The model assumes that cross-immunity depends on the set of strains seen in prior infections but not on the order of such infections. Furthermore, we exclude the possibility of co-infection or superinfection (May & Nowak 1994, 1995; Nowak & May 1994).

Let the set $H = \{1,2,3,...,n\}$ label the n-strains present in the system. The 2^n subsets of H span all possible unordered histories of strain exposure. We define $S_{\mathcal{J}}$ as the proportion of the population that are currently uninfected and have prior infection history $\mathcal{J} \subseteq H$, and $I^i_{\mathcal{J}}$ as the proportion currently infected by strain $i \in H$ and with infection history $\mathcal{J}$, where $i \notin \mathcal{J}$. This formulation requires $2^n + n2^{n-1}$ state variables to describe a system with n strains. System dynamics are defined by the following differential equations:

$$\begin{aligned}
\dot{S}_{\varnothing} &= \mu N - \mu S_{\varnothing} - \sum_{i \in H} \Lambda^i S_{\varnothing},\\
\dot{S}_{\mathcal{J}} &= v \sum_{j \in \mathcal{J}} I^j_{\mathcal{J}\setminus j} - \mu S_{\mathcal{J}} - \sum_{i \notin \mathcal{J}} \sigma^i_{\mathcal{J}} \Lambda^i S_{\mathcal{J}},\\
\dot{I}^i_{\mathcal{J}} &= \sigma^i_{\mathcal{J}} \Lambda^i S_{\mathcal{J}} - \alpha I^i_{\mathcal{J}},
\end{aligned} \tag{2.1}$$

where N is the equilibrium population size, μ is the per capita death and birth rate, α is the rate of loss of infectiousness (sum of death and recovery rates), $\sigma^i_{\mathcal{J}}$ is the reduction in susceptibility to strain i induced by infection history $\mathcal{J}$, and β_i is the infectiousness of strain i. The force of infection for strain i is given by $\Lambda^i = \beta_i \sum_{M \subseteq H/i} I^i_M$, where $H\setminus i$ stands for all elements in H with the exception of strain i. Note that $\sigma = 0$ corresponds to total cross-immunity while $\sigma = 1$ corresponds to no cross-immunity.

Solution of the general form of this model at an arbitrary number of strains appears intractable. We therefore simplify the problem to examine only the symmetric multi-strain system where all strains share the same epidemiological properties. At the equilibrium of such a symmetric system, $I^k_L = I^i_{\mathcal{J}} \equiv I_m$ and $S_L = S_{\mathcal{J}} \equiv S_m$ for any strains k and i and L or $\mathcal{J} \subseteq H$ provided the number of elements (m) in L and $\mathcal{J}$ are the same. For such a system, $\beta_i = \beta$, and $\sigma^i_{\mathcal{J}} = \sigma_m$ where the subscript m is the number of previous infections experienced. Moreover, $\Lambda^i = \Lambda = \beta y_{\mathrm{T}}/n$, where y_{T} is the total proportion of infected population (prevalence) in the system.

Using standard algebraic techniques, we can express the equilibrium equations as a recursion relation $b_m I_m - a_m I_{m-1} = 0$ subject to the constraint

$$y_{\mathrm{T}} = \sum_{m=0}^{n-1} \frac{n!}{(n-m-1)!m!} I_m,$$

where a_m and b_m are simple functions of y_{T}. The system can be formally solved using standard difference equations techniques to obtain a master equation

$$y_{\mathrm{T}} = e \sum_{m=0}^{n-1} (1-e)^m \prod_{i=0}^{m} (1 - 1/[1 + R_0 \sigma_i y_{\mathrm{T}}(1 - i/n)/e]), \tag{2.2}$$

where $e = \mu/\alpha$ and the basic reproduction number for any strain (Anderson & May 1991) $R_0 = \beta/\alpha$. This is a polynomial of degree n, but one can prove that there is a single positive definite, and thus biologically meaningful, solution.

One can exactly solve the equilibrium for some special cases. When cross-immunity is total (i.e. $\sigma_i = 0$) then $y_{\mathrm{T}} = e(1 - 1/R_0)$: the result obtained for the one-strain system, but with these infections being evenly distributed across all n strains. As $n \to \infty$ and for $\sigma_m = \sigma$ for all m (constant cross-immunity profile), it can be shown that

$$y_{\mathrm{T}}(n \to \infty) = \frac{\sigma(R_0 - e) - 1 + \sqrt{(\sigma(R_0 - e) - 1)^2 + 4e\sigma(R_0 - 1)}}{2\sigma R_0}, \tag{2.3}$$

which in the limit of no cross-immunity ($\sigma_i = 1$) has the simple form $y_{\mathrm{T}} = 1 - 1/R_0$.

There are also an important set of approximate solutions that may be derived. The most useful of these is based on the idea that the dynamics of a single strain in the presence of $n - 1$ other strains can be mapped onto those of a single strain on its own through a parameter transformation. In deriving this mapping we want to reproduce the two key epidemiological quantities of the exact system: total infection prevalence and incidence. We model the effect of cross-immunity generated by other strains on the transmission of one strain as an increased rate at which individuals leave the susceptible class of that strain; i.e. an increased death rate, $\mu_{\mathrm{eff}} = \phi\mu$, where $\phi > 1$. However, by changing the death rate, we also change the net rate at which individuals enter the susceptible population, μN. As this is unaffected by cross-immunity, we therefore also need to transform the equilibrium population size $N \to N_{\mathrm{eff}} = N/\phi$. Thus the one-strain system with dynamics equivalent to the multi-strain system has a

smaller total population size. The rate of loss of infectiousness, α, is unchanged by the effect of cross-immunity. However, for infection incidence to be identical in both the full and single-strain systems, we require $\beta_{\text{eff}} I/N_{\text{eff}} = \beta I/N$ (given that the number of infected individuals, I, in both systems is identical), implying $\beta_{\text{eff}} = \beta/\phi$, or equivalently, $R_{0\text{eff}} = R_0/\phi$. We are then left with determining the value of ϕ, which ensures that I is identical for both systems. This requires the following equation to be solved for ϕ:

$$y_{\text{T}} = ne\left(1 - \frac{1}{R_0/\phi}\right). \tag{2.4}$$

With knowledge of the exact solution to equation (2.2), this expression is useful on its own when examining the stochastic dynamics of the multi-strain system (see below). However, in the case of constant cross-immunity we can directly write down an approximate expression for $R_{0\text{eff}}$:

$$1/\phi = S_0 + \sigma(1 - S_0 - (n-1)y_{\text{T}}/n), \tag{2.5}$$

where S_0 is the proportion of individuals susceptible to all strains in the $n-1$ strain system, approximately given by

$$S_0 = 1/[1 + R_0(n-1)y_{\text{T}}/ne]. \tag{2.6}$$

Combining equations (2.4)–(2.6) gives a cubic, but owing to the complexity of the relevant root we omit the closed form here for simplicity, though it can be shown that this approximation is excellent when cross-immunity is intense ($\sigma \ll 1/R_0$).

Conversely, for large σ (more precisely, for $\sigma \gg 1/R_0$) it is possible to demonstrate that a good approximation is

$$y_{\text{T}}(n;\sigma) = (1 - (1-e)^n)\left(1 - \frac{1}{\sigma R_0}\right), \tag{2.7}$$

which reduces to approximately $ne(1 - 1/\sigma R_0)$ for small e and n (i.e. $\phi \approx 1/\sigma$).

In general, however, the equilibrium equation is analytically intractable, but is numerically soluble by successive approximations using Banach's fixed point theorem, with the solution being the fixed point of the sequence generated by the right-hand side of the master equation. The numerical solution converges swiftly to the exact solution with the computational time increasing linearly as a function n.

Figure 1 displays the total fraction of infected population y_{T} as a function of the number of strains and the cross-immunity profile σ_m for a constant cross-immunity profile $\sigma_m = \sigma$ and for a geometric profile $\sigma_m = \sigma^m$. For the geometric profile, σ has a simple interpretation as the factor by which individual susceptibility decreases after each additional infection.

For the constant cross-immunity profile (figure 1*a*), y_{T} matches the large-σ analytical approximation given above to a high degree of accuracy when $\sigma \gg 1/R_0$, saturation being reached by $n \sim 10^4$ with $y_{\text{T}} = 1 - 1/\sigma R_0$. For σ closer to $1/R_0$, the approximation is less adequate and very different behaviour is seen for $\sigma < 1/R_0$, with no region of linear increase with n and a much lower saturation level of y_{T} (discussed further below). In this regime the small-σ approximation given above is accurate.

For the geometric cross-immunity profile (figure 1*b*), the initially linear increase in y_{T} leading to saturation is also observed for $\sigma \gg 1/R_0$, but y_{T} saturates at a much reduced level. This reflects the fact that this cross-immunity much more effectively prevents hosts from getting infected by more than few strains over their lifetime.

We find that the behaviour versus R_0 is qualitatively similar to the behaviour versus σ, indicating a degree of dynamical equivalence between these two parameters. More insight into this aspect of the dynamics can be seen in the sharp transition occurring at $R_0\sigma \approx 1$[1] evident in figure 1*a*. The transition precisely occurs at $(R_0 - e)\sigma = 1$, but since $R_0 \gg e$ in typical multistrain systems, $R_0\sigma \approx 1$ is a highly accurate approximation. In the case of almost total cross-immunity ($\sigma \sim 0$), very intense inter-strain competition leads to small y_{T} irrespective of the number of strains, and all strains share approximately the same force of infection as in the one-strain system. With increasing σ, the behaviour of the constant cross-immunity profile system (figure 1*a*) diverges from that of the geometric profile (figure 1*b*). In the former case, a bifurcation occurs at $\sigma \approx 1/R_0$, above which y_{T} grows rapidly, only eventually saturating because, of susceptible depletion. This transition arises because, for $\sigma < 1/R_0$, the effective reproduction number for any strain spreading in the proportion of the population with prior infection, $R_0\sigma$, is below 1, preventing self-sustaining transmission occurring in that sub-population. The same abrupt transition is not observed for a geometric cross-immunity profile, where y_{T} grows slowly with increasing σ until $\sigma = 1$. In the case of constant cross-immunity, once $R_0\sigma \approx 1$ is crossed every partly susceptible sub-population (S_m) in the system can support self-sustaining transmission. In the case of geometric immunity, crossing the $R_0\sigma \approx 1$ threshold only allows the S_1 populations to support self-sustaining infection with S_2 and higher populations still having an effective reproduction number $R_0\sigma^m$ below 1. Hence for constant cross-immunity, the abrupt transition seen reflects a major system restructuring where all state variables become appreciable at the same time, while for geometric cross-immunity the S_m state variables become appreciable only as the effective reproduction number $R_0\sigma^m$ exceeds 1 for the respective sub-population S_m.

3. AVERAGE NUMBER OF STRAINS IN THE PRESENCE OF MUTATIONS AND EXTINCTIONS

Section 2 considers the equilibrium properties of multi-strain pathogen systems with a fixed number of strains. In reality, the number of strains in any multi-strain system is susceptible to fluctuations arising from the generation of new strains through mutation, and stochastic extinction of strains driven by competition and finite host population size. However, direct analysis (or even simulation) of the fully stochastic multi-strain SIR model for an arbitrary number of strains is intractable, so here we use the approximation given above to map the dynamics of a single strain co-circulating with $n-1$ other strains onto the simple (single-strain) stochastic SIR model.

Assuming that each strain generated via mutation is as antigenically distant from pre-existing strains as pre-existing strains are from each other, and that the total mutation rate is proportional to the number of hosts infected, then the rate at which new strains are generated and reach equilibrium levels can be approximated by $\varepsilon(1 - 1/R_{0\text{eff}})Ny_{\text{T}}$. Here ε is the mutation rate per infected individual and $1 - 1/R_{0\text{eff}}$ is the probability that a mutation

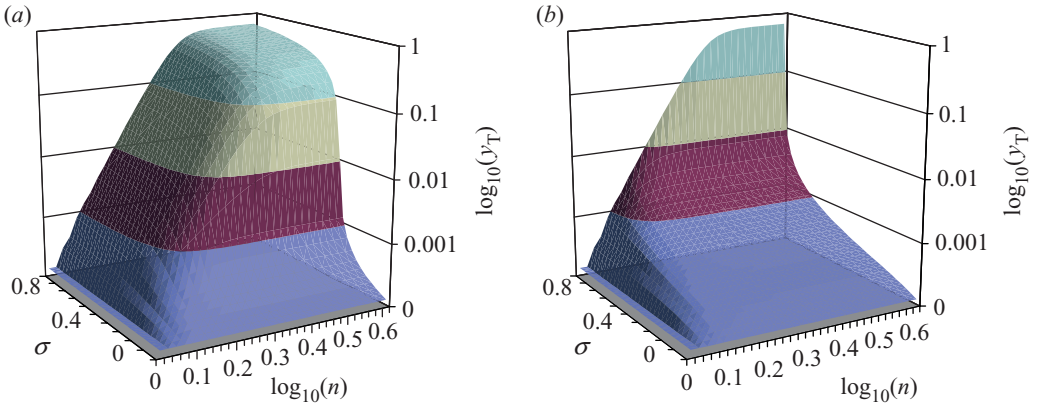

Figure 1. Total infection prevalence y_T, at symmetric equilibrium of a multi-strain system for (*a*) constant ($\sigma_i = \sigma$) cross-immunity profile as a function of the number of strains n and σ for $R_0 = 3$ and $e = 0.0002$ (corresponding to a recovery period $\nu^{-1} = 4$ days and a host lifetime $\mu^{-1} = 50$ years). (*b*) As (*a*) but with geometric ($\sigma_i = \sigma^i$) cross-immunity profile. For constant cross-immunity, the figure demonstrates the existence of three distinct dynamical regimes in multi-strain systems: low prevalence at $R_0\sigma < 1$, rapid growth in prevalence at $R_0\sigma \approx 1$, and prevalence saturation for $R_0\sigma \gg 1$. For geometric cross-immunity, the $R_0\sigma \approx 1$ transition is less dramatic and the system saturates swiftly at lower prevalence levels.

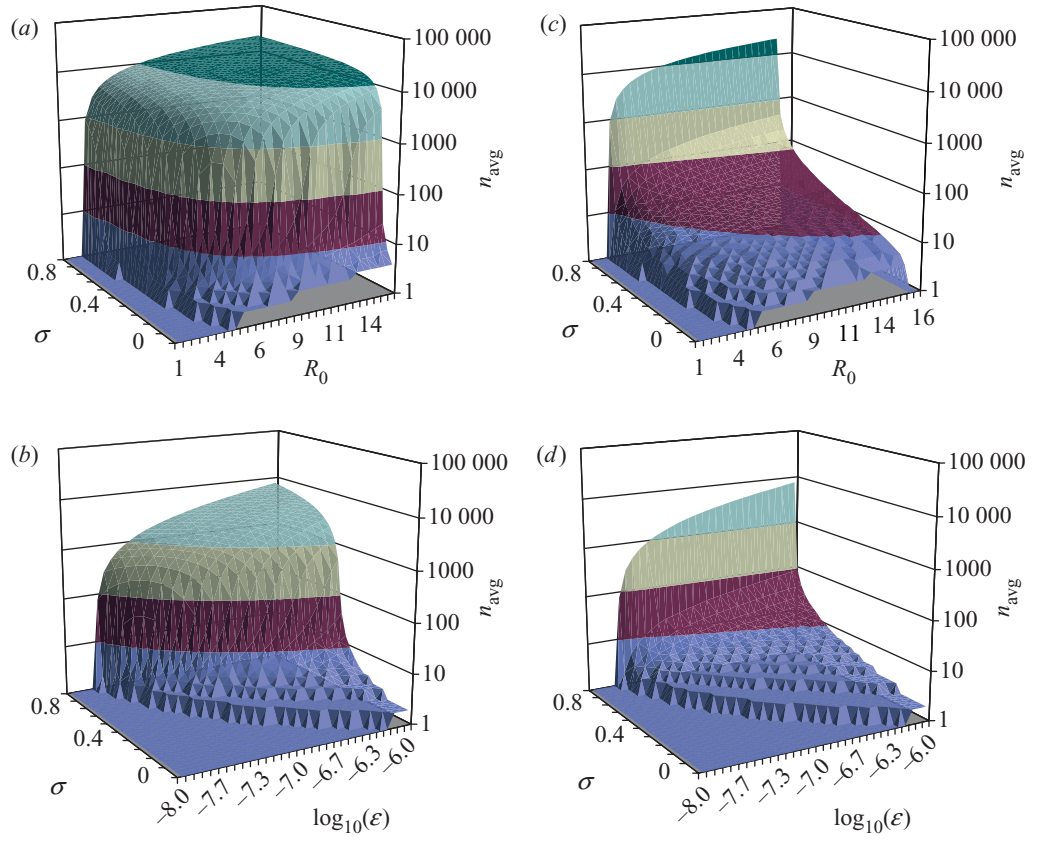

Figure 2. Strain diversity in finite host populations. The average number of strains in the system (n_{avg}) for constant cross-immunity profile as a function of: (*a*) σ and R_0; and (*b*) σ and ε. (*c*) and (*d*) As (*a*) and (*b*) respectively but with a geometric $\sigma_i = \sigma^i$ cross-immunity profile. For constant cross-immunity, the figure illustrates the existence of three distinct dynamical regimes in multi-strain systems: low or no diversity at $R_0\sigma < 1$, rapid growth in diversity at the diversity threshold $R_0\sigma \approx 1$, and moderate growth in diversity for $R_0\sigma \gg 1$. For geometric immunity, the $R_0\sigma \approx 1$ transition is smoother. For all plots $N = 10^8$ and $\varepsilon = 10^{-7}$, with other unvaried parameters as in figure 1.

causing a single infection with a new strain will generate a substantial epidemic (Bailey 1975), where $R_{0\text{eff}}$ is the effective reproduction number of a new strain in the presence of $n-1$ other strains (as given by equation (2.4)). The total size of the infected population in the system is Ny_T.

Stochastic extinction is incorporated into the model using the approximate analytical result on the mean time to extinction of a single strain derived by Nasell (1999), and the mapping of the multi-strain SIR system onto the simple SIR model discussed above. We modify the expression for

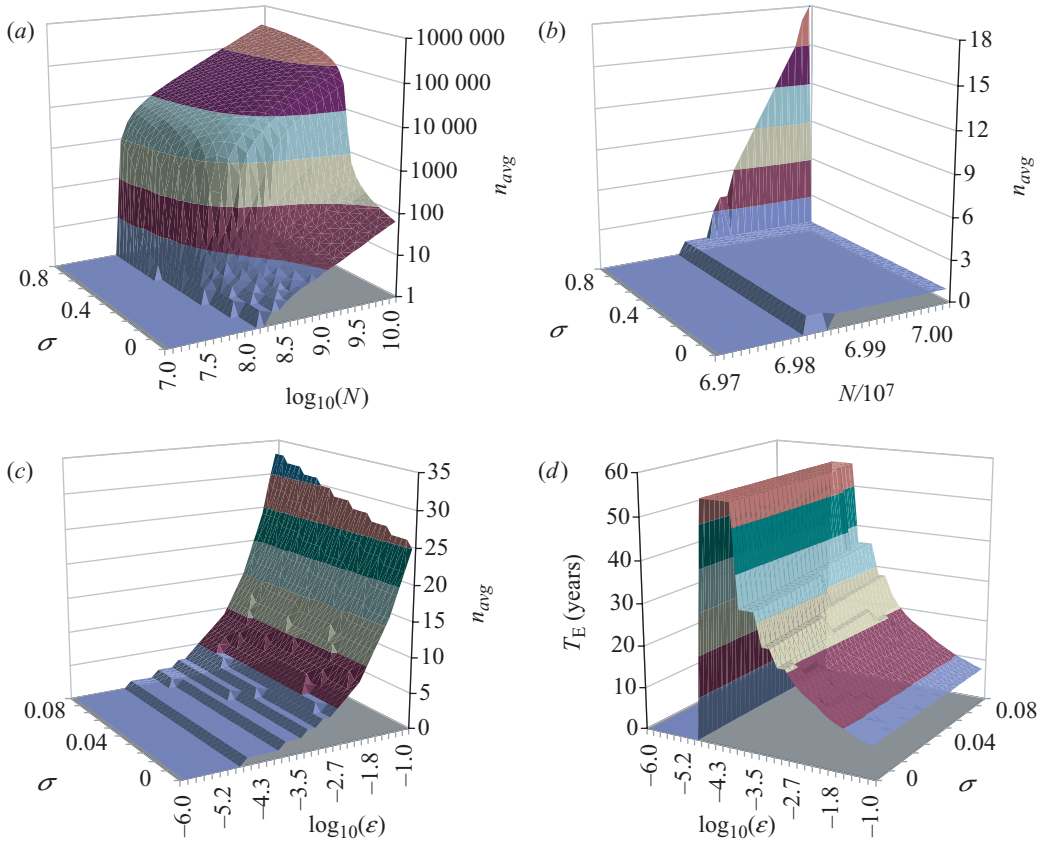

Figure 3. Strain diversity as a function of host population size and stability of one-strain systems. (*a*) The average number of strains (n_{avg}) as a function of host population size N and cross-immunity σ for constant cross-immunity at a mutation rate of $\varepsilon = 10^{-7}$, a reproductive number of $R_0 = 3$, a recovery period $\nu^{-1} = 4$ days and a host lifetime $\mu^{-1} = 50$ years. (*b*) As (*a*) but near the single-strain persistence threshold. Three distinct types of dynamics are seen near the threshold: total strain extinction below the critical community size, a stable but dynamical multi-strain system for large N and σ, and a stable and largely static regime where there is only one strain in the system. (*c*) The average number of strains n_{avg} as a function of the mutation rate ε and cross-immunity σ for a system with population size $N = 10^7$ and geometric immunity profile (results similar for constant profile). Other parameters as (*a*). (*d*) As (*c*) but plotting the time to extinction T_E for each strain as a function of ε and σ. Small values of T_E indicate that all strains face eventual extinction.

the time to extinction, T_E, in the single-strain SIR model to account for the fact that the death rate, host population size, and reproduction number in the single-strain system are transformed in the n-strain environment as a result of strain competition:

$$T_E = \frac{\sqrt{2\pi\mu_{eff}^2(R_{0eff}-1)N_{eff}/\alpha^2}}{\mu_{eff}R_{0eff}}\exp\left(\frac{\mu_{eff}^2(R_{0eff}-1)N_{eff}}{2\alpha^2}\right). \tag{3.1}$$

We thereby arrive at a rate of change for the number of strains in a host population of size N of

$$\frac{dn}{dt} = \varepsilon(1-1/R_{0eff})Ny_T - \frac{n}{T_E(\mu_{eff},N_{eff},R_{0eff})}. \tag{3.2}$$

The average number of strains a population of size N can support is given by the equilibrium value of n ($dn/dt = 0$).

Figures 2 and 3 show dependence of the average number of strains in the system (n_{avg}) on cross-immunity and each of the parameters R_0, ε and N. For constant cross-immunity a clear diversity threshold for n_{avg} is seen, driven by the deterministic $R_0\sigma \approx 1$ phase transition discussed previously. However, while for large host population sizes the transition is at $R_0\sigma \approx 1$, for smaller populations it moves above $R_0\sigma \approx 1$ because in a finite-sized system a secondary condition also needs to be satisfied: that the host population exceeds the critical community size for single- or multiple-strain persistence. When $R_0\sigma \gg 1$, strain diversity increases rapidly with increasing R_0, ε, σ, N and e. In this regime, the dynamics are largely determined by the prohibition of co-infection implicit in the model, which lowers per-strain infection prevalence for high n and hence increases strain extinction rates.

The $R_0\sigma < 1$ low-prevalence regime can support much more limited diversity, though this can be maximized by high R_0, large N, a long infectious period (e) or a high mutation rate (ε). Diversity increases fastest with increasing N, because increases in N increase both the strain generation rate and lower the extinction rate. Increases in other parameters have a less dramatic impact on diversity, because in the asymptotic limit they affect only the strain generation rate (ε) or extinction rate (e and R_0). However, it can be shown that the rate of growth in diversity as a function of e ($T_E \sim e\exp(e^2)$) is much faster than that with R_0 ($T_E \sim \sqrt{(R_0-1)}/R_0\exp(R_0-1)$).

Figure 2*c*,*d* shows how cross-immunity, which increases geometrically with the number of infections experienced, greatly limits diversity compared with the constant cross-immunity profile results shown in figure 2*a*,*b*. This is

because geometrically increasing cross-immunity limits the number of infections a host may experience over its lifetime. As discussed above, the $R_0\sigma \approx 1$ threshold for a geometric immunity profile does not generate the major system restructuring caused in the constant-immunity case. Hence antigenic diversity depends critically on how cross-immunity accrues with multiple infections.

The dependence of strain diversity on host population size (figure 3*a*,*b*) merits attention. The concept of the critical community size—the host population size below which disease extinction by random chance is likely—is well established for single-strain pathogens (Bartlett 1960; Black 1966). We extend this concept to define the quantity $N_{\mathrm{crit}}(n)$: the minimum population size for which long-term persistence of n strains is possible. Figure 3*a* shows that for $N \gg N_{\mathrm{crit}}(1)$ and either $R_0\sigma \ll 1$ or $R_0\sigma \gg 1$, diversity increases linearly with population size, *i.e.* $N_{\mathrm{crit}}(n) \sim n$.

A more detailed look at the behaviour of equilibrium diversity near the single-strain persistence threshold is informative (figure 3*b*). Three dynamical regimes can be discerned: no strains below the critical community size, multiple strains for large N and σ, and an intermediate region where there is only one strain in the system. These results highlight the fundamental dynamical distinction between one- and multi-strain systems, namely that cross-immunity—which increases extinction rates—affects only systems with more than one strain.

These results demonstrate how emerging pathogens have three fates: extinction, single-strain endemicity, and multi-strain endemicity. For a pathogen with epidemiological properties, putting it at the interface between single- and multi-strain systems, small evolutionary changes (or increases in host population size) can rapidly move a disease from single-strain to multiple-strain endemicity. Multi-strain endemicity does not necessarily imply long-term persistence of all strains, but rather that there is enough strain generation to balance strain extinction, leading to a system with continuous strain turnover. The time to extinction for individual strains in multi-strain systems is typically much lower than for single-strain systems, and for diseases such as influenza (where individual strains survive for at most a few years), it is orders of magnitude lower.

Figure 3*c*,*d* focuses on this type of dynamical regime, and plots the average number of strains and the time to extinction for a single strain as a function of cross-immunity and mutation rate for a pathogen with typical influenza-like parameters. It is important to note that the limited diversity of influenza A observed at any point in time implies that this pathogen must fall below the $R_0\sigma \approx 1$ diversity threshold. The observed rate of strain emergence is generated primarily by the large size of the host population and the high mutation rate, and diversity is restricted by intense cross-immunity between strains.

Short-term non-specific immunity has recently been proposed as being key to explaining the limited diversity of influenza (Ferguson *et al.* 2003). In the current framework, non-specific immunity would add a time delay between recovery from infection and re-entry into the susceptible population for infection by other strains. Consequently such immunity lowers strain diversity and improves model realism by producing reasonable levels of strain diversity at population sizes considerably larger than the relatively small value of 10^7 used in figure 3*b*. Overall the pattern of strain replacement dynamics generated by our simple model is broadly consistent with influenza A evolution. However, it is arguable that the simple cross-immunity structure used here (where all strains compete equally with each other) makes reproducing the qualitative pattern of influenza evolution artificially easy compared with more realistic frameworks (Ferguson *et al.* 2003) which permit strains to progressively escape pre-existing host immunity through mutation.

4. CONCLUSIONS

The goals of this work differed somewhat from previous theoretical analyses of multi-strain infectious disease systems: rather than focusing on characterizing the impact of cross-immunity on transmission dynamics for systems with a fixed (usually small) number of strains (Gupta *et al.* 1998), here we have examined how system properties change as a function of the number of strains. Moreover, this paper presents one of the first analyses of how the interactions between diversification processes (i.e. mutation), inter-strain competition (via cross-immunity) and demographic stochasticity determine the equilibrium diversity of a pathogen in a finite-sized host population. We solved a complex multi-strain model at equilibrium for an arbitrary number of strains and for a general cross-immunity profile. The solution derived permits easy numerical evaluation, but we also derived accurate analytical approximations to the exact solution valid in a variety of parameter regimes. Genetic variability and stochastic extinction were incorporated into the modelling framework through the derivation of a novel approximate mapping of multi-strain systems onto the stochastic single-strain SIR model.

It is interesting to compare our approach with status-based models (Gog & Grenfell 2002) in which cross-immunity is polarized, acts on infectiousness, and does not incorporate accumulation of immune memory. Status-based models give a very rapid decay in the fraction of infected population as a function of the number of strains and cross-immunity, with the effect of cross-immunity being much more severe compared with immune-memory-based approaches of the type used here. When cross-immunity is completely effective, both approaches reduce to a one-strain SIR model. For zero cross-immunity, status-based models yield n independent and uncoupled SIR systems while in our model, even though no cross-immunity is present, strains do interact because of prohibition of co-infection. Overall results from the two approaches converge for small number of strains but diverge substantially for large numbers of strains. This discrepancy makes model dependence a cause for concern. A detailed clarification of the extent of homology in system dynamics between these approaches would therefore be valuable.

This work has identified three distinct dynamical trajectories for novel genetically variable pathogens entering a naive host population: inevitable extinction as a result of low host population size and/or transmissibility, stable and largely static single-strain population structure, and multiple-strain dynamics with stable diversity but possibly high rates of strain turnover. Which outcome is realized depends critically on the reproduction number of the pathogen, the intensity of cross-immunity and, to a lesser extent, on the

rate of mutation and duration of infectiousness. This delineation of outcomes results from the three distinct dynamical regimes exhibited by multi-strain systems at equilibrium: a region below the $R_0\sigma \approx 1$ transition marked by low prevalence and no or limited diversity, the $R_0\sigma \approx 1$ diversity threshold characterized by explosive increase in disease prevalence and strain diversity, and last, a region with saturating growth in prevalence and moderate increase in diversity above $R_0\sigma \approx 1$. The severity of the $R_0\sigma \approx 1$ transition is determined by the manner in which immunity accumulates with repeated strain infections. Progressively (e.g. geometrically) increasing immunity gives a smooth transition while constant immunity gives a sudden phase transition and a near total restructuring of the system at $R_0\sigma \approx 1$.

A novel aspect of the work presented here is the inclusion of demographic stochasticity and hence the insight given into how the diversity of pathogen populations depends on host population size. Our results indicate that outside the threshold regime of $R_0\sigma \approx 1$, the diversity of pathogen populations is expected to depend linearly on host population size for a homogenously mixing population, agreeing with the population genetics result that diversity of clonal organisms increases linearly with effective population size, given that effective population size is proportional to host population size for directly transmitted pathogens. However, our analysis extends simple population genetic analyses by giving quantitative insight on how epidemiological parameters (such as transmissibility, cross-immunity and infectious period), as well as host population size, determine pathogen diversity. As such, the methods presented may be of use in improving assessments of the likely potential consequence of increasing global population size on pathogen diversity. Another interesting though technically challenging future extension of the work presented here would be to explore how host metapopulation structure might affect the relationship between pathogen diversity and host population size.

Simplifications were necessary to examine the properties of multi-strain systems with arbitrary diversity. In particular, the current analysis is restricted to examining equilibrium dynamics for the completely symmetric system (i.e. all strains were assumed to share identical epidemiological properties). In extending the deterministic analysis to incorporate stochastic extinction, we therefore implicitly assumed highly simplified strain establishment dynamics: new strains arising through mutation either go extinct with probability $1/R_{0\mathrm{eff}}$ or immediately reach equilibrium prevalence. Because influenza strains in reality are unlikely to reach equilibrium, this assumption (together with that of symmetric cross-immunity between all strains) may in part explain the otherwise encouraging ability of this framework to reproduce influenza-like dynamics (i.e. low average diversity but a high rate of strain turnover) without recourse to the additional mechanisms necessary in more realistic models (Ferguson *et al.* 2003).

Ongoing work is therefore focusing on relaxing some of these simplifications and examining systems with randomly distributed R_0 and inter-strain cross-immunity. The challenges are considerable, in particular because incorporation of such heterogeneity disrupts the interior equilibrium studied here, with more extreme parameter choices already shown to give rise to limit cycles and sometimes chaotic behaviour (Gupta *et al.* 1998; Lin *et al.* 1999). However, characterizing the statistical dynamics of large numbers of strains with heterogeneous but potentially strong immunity-mediated interactions is key to gaining a more complete understanding of both the epidemiological and evolutionary dynamics of genetically variable pathogens.

L.J.A.-R. and N.M.F. thank the Medical Research Council, and N.M.F. also thanks The Royal Society and the Howard Hughes Medical Institute for research funding.

REFERENCES

Ackerman, E., Longini, J. I. M., Seaholm, S. & Hedin, A. S. 1990 Simulation model for viral interference in influenza. *Int. J. Epidemiol.* **19**, 444–454.

Anderson, R. M. & May, R. M. 1991 *Infectious diseases of humans: dynamics and control*. Oxford University Press.

Andreasen, V., Levin, S. A. & Lin, J. 1996 A model of influenza A drift evolution. *Z. Angew. Math. Mech.* **52**, 421–424.

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

Bailey, N. T. J. 1975 *The mathematical theory of infectious diseases and its applications*. London: C. Griffin and Co. Ltd.

Bartlett, M. S. 1960 The critical community size for measles in the US. *J. R. Stat. Soc.* A **123**, 37–44.

Black, F. 1966 Measles endemicity in insular populations: critical community size and its evolutionary implication. *J. Theor. Biol.* **11**, 207–211.

Castillo-Chavez, C., Hethcote, H. W., Andreasen, V., Levin, S. A. & Liu, M. W. 1989 Epidemiological models with age structure, proportionate mixing, and cross-immunity. *J. Math. Biol.* **27**, 233–258.

Cooper, B. S. 2001 Pathogen population dynamics: the age of the strain. *Trends Microbiol.* **9**, 199–200.

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

Ferguson, N. M. & Andreasen, V. 2002 The influence of different forms of cross-protective immunity on the population dynamics of antigenically diverse pathogens. In *Mathematical approaches for emerging and re-emerging infectious diseases: models, methods and theory,* vol. 125 (ed. S. M. Blower & C. Castillo-Chavez), pp. 157–169. New York: Springer.

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

Gog, J. R. & Grenfell, B. T. 2002 Dynamics and selection of many-strain pathogens. *Proc. Natl Acad. Sci. USA* **99**, 17 209–17 214.

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

Gomes, M. G. M. & Medley, G. F. 2002 Dynamics of multiple strains of infectious agents coupled by cross-immunity: a comparison of models. In *Mathematical approaches for emerging and re-emerging infectious diseases: models, methods and theory*, vol. 125 (ed. S. M. Blower & C. Castillo-Chavez), pp. 171–191. New York: Springer.

Grenfell, B. & Gog, J. 2001 Pathogen strains: no joke. *Trends Ecol. Evol.* **16**, 272.

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**, 327–332.

Gupta, S. & Maiden, M. C. 2001 Exploring the evolution of diversity in pathogen populations. *Trends Microbiol.* **9**, 181–185.

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

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

Gupta, S., Ferguson, N. & Anderson, R. 1998 Chaos, persistence and the evolution of strain structure in populations of antigenically variable infectious agents. *Science* **240**, 912–915.

Lin, J., Andreasen, V. & Levin, S. A. 1999 Dynamics of influenza A drift: the linear three-strain model. *Math. Biosci.* **162**, 33–51.

May, R. M. & Nowak, M. 1994 Superinfection, metapopulation dynamics, and the evolution of diversity. *J. Theor. Biol.* **170**, 95–114.

May, R. M. & Nowak, M. A. 1995 Coinfection and the evolution of parasite virulence. *Proc. R. Soc. Lond.* B **261**, 209–215.

Nasell, I. 1999 On the time to extinction in recurrent epidemics. *J. R. Stat. Soc.* B **61**, 309–330.

Nowak, M. & May, R. M. 1994 Superinfection and the evolution of parasite virulence. *Proc. R. Soc. Lond.* B **255**, 81–89.

Simonsen, L., Clarke, M. J., Schonberger, L. B., Arden, N. H., Cox, N. J. & Fukuda, K. 1998 Pandemic versus epidemic influenza mortality: a pattern of changing age distribution. *J. Infect. Dis.* **178**, 53–60.

WHO/UNICEF 2003 *Africa malaria report*. WHO/UNICEF.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.