Incorporating demographic stochasticity into multi-strain epidemic models: application to influenza A

Pavlo Minayev* and Neil Ferguson

MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College of Science, Technology & Medicine, Norfolk Place, London W2 1PG, UK

We develop mathematical models of the transmission and evolution of multi-strain pathogens that incorporate strain extinction and the stochastic generation of new strains via mutation. The dynamics resulting from these models is then examined with the applied aim of understanding the mechanisms underpinning the evolution and dynamics of rapidly mutating pathogens, such as human influenza viruses. Our approach, while analytically relatively simple, gives results that are qualitatively similar to those obtained from much more complex individually based simulation models. We examine strain dynamics as a function of cross-immunity and key transmission parameters, and show that introducing strain extinction and modelling mutation as a stochastic process significantly changes the model dynamics, leading to lower strain diversity, reduced infection prevalence and shorter strain lifetimes. Finally, we incorporate transient strain-transcending immunity in the model and demonstrate that it reduces strain diversity further, giving patterns of sequential strain replacement similar to that seen in human influenza A viruses.

Keywords: multi-strain model; population dynamics; transient strain-transcending immunity; cross-immunity; strain diversity; influenza dynamics and evolution

1. INTRODUCTION

Modelling of the transmission of multi-strain pathogens to date has either used compartmental deterministic frameworks (Andreasen et al. 1997; Gupta et al. 1998; Gog & Grenfell 2002) or large-scale individually based simulations (Ferguson et al. 2003; Koelle et al. 2006). While the former retain sufficient simplicity to allow (quasi-) analytical insight into system dynamics, they cannot satisfactorily capture the intrinsically stochastic nature of strain creation (via mutation or recombination) and extinction (Minayev & Ferguson 2008). Simulations can capture such processes more realistically, and past work has shown that stochasticity plays a key role in explaining the evolution of influenza A in humans (Ferguson et al. 2003; Koelle et al. 2006), but in this case structural and computational complexity means that insight into key drivers of the dynamics can be compromised. Here we try to bridge that gap, by introducing discrete strain extinction and generation (i.e. mutation) processes into previously developed deterministic model frameworks (Minayev & Ferguson 2008). Our aim is to find the minimally complex model capable of giving an at least qualitatively realistic description of the evolutionary dynamics of influenza, i.e. a model giving relatively low standing diversity of extant strains despite a very large potential diversity, together with frequent sequential strain replacement (Bush et al. 1999; Smith et al. 2004).

Deterministic epidemic models governed by a set of nonlinear differential equations have the well-recognized limitation that, as all the dynamical variables are real valued, a strain, once it becomes established, never goes completely extinct, and instead continues to infect a (perhaps vanishingly) small proportion of the population. As a result, very long-period epidemic cycles of individual strains can occur, despite trough prevalences of $10^{-20}$ or less. Equilibrium strain diversity can therefore be unrealistically high. In order to provide better quantitative estimates for disease persistence or elimination, more biologically realistic models need to take into account the probability of extinction of the disease as soon as the proportion of infections falls to low levels. We therefore extend the fully deterministic compartmental model explored in Minayev & Ferguson (2008) by introducing an extinction threshold according to the following algorithm: if there is only one individual infected with a particular strain in the population and no transmission takes place during the mean infectious period, this strain is eliminated from the simulation (i.e. the
proportion infectious with this strain is set to 0). This modification of course introduces the notion of population size into the model.

Another factor that is important to a realistic description of the process of emergence of new strains is the discrete and stochastic nature of mutation. Therefore, we model mutation stochastically, by picking the number of individuals who experience mutations during their infectious period from a Poisson distribution. We assume, as we did for the deterministic model, that mutation only causes a one allele change and that after mutation hosts become infectious with the mutated strain.

We first develop the model, before exploring how its dynamics differ from the deterministic model without extinction or stochastic mutation. We study model variants with and without transient strain-transcending immunity of the type first postulated in Gupta et al. (1998) to explore to what extent such immunity is necessary to reproduce influenza-like evolutionary dynamics.

2. MODEL DEFINITION

We consider pathogen strains to be sequences of $N_L$ antigenic loci with each locus able to be occupied by one of $N_k$ alternative alleles, giving a total of $N_k^N_L$ different combinations. Infection, recovery from infection, birth and death are described deterministically, and are governed by the same equations as in our previous work (Minayev & Ferguson 2008). Namely, the dynamics of the population susceptible to strain $i$ is described by

$$\frac{ds_i}{dt} = - (\lambda_i + \mu) s_i + \mu,$$  \hspace{1cm} (2.1)

where $\lambda_i = \beta y_i$ is the force of infection of strain $i$; $\mu$ is the birth and death rate (births balance deaths); $\beta$ is the transmission coefficient; and $y_i$ is the proportion of the population infectious with strain $i$.

The dynamics of $s_i^{(k)}$, the proportion of the population susceptible to any strain that shares alleles with $i$ but contains not more than $k$ alleles different from $i$, is given by

$$\frac{ds_i^{(k)}}{dt} = \mu (1 - s_i^{(k)}) - s_i^{(k)} \sum_{j \neq i, d_{ij} \leq k} \lambda_j.$$  \hspace{1cm} (2.2)

In the sum in (2.2), only those strains $j$ whose genetic distance from strain $i$ does not exceed $k$ are considered. Index $k$ ranges from 1 to $(N_L - 1)$ and by definition we put $s_i^{(N_L)} = 0$ and $s_i^{(0)} = s_i$.

We assume, as we did in Minayev & Ferguson (2008), that, upon infection with strain $i$, a host gains a partial protection against any strain $j$. Cross-immunity acts to reduce the transmission success by a factor $(1 - \gamma(d_{ij}))$, where $\gamma(d_{ij})$ is the cross-immunity function that we suppose linearly decreases with the genetic (Hamming) distance $d_{ij}$ between strains $i$ and $j$.

$$\gamma(d) = \begin{cases} 
1, & \text{if } d = 0, \\
b \cdot [1 - (d - 1)a], & \text{if } d = 1, 2, \ldots \leq (1/a + 1) < N_L, \\
0, & \text{otherwise},
\end{cases}$$  \hspace{1cm} (2.3)

For a more detailed discussion of the deterministic equations (2.1) and (2.2) and the function (2.3) describing the cross-immune response, the interested reader can see our previous paper Minayev & Ferguson (2008).

The dynamics of the infectious compartments is governed by

$$\frac{dy_i}{dt} = \lambda_i \sum_{j=0, j \neq i}^{(N_L-1)} (s_i^{(j)} - s_i^{(j+1)}) (1 - \gamma(j + 1)) - \sigma y_i + SMT,$$  \hspace{1cm} (2.4)

where $\sigma$ is the recovery rate and SMT is a stochastic (Poisson) term representing the mutation process. Mutation is modelled at the level of individual hosts and we assume that, as a result of mutation, a host becomes infectious with a new strain. The rate of the corresponding Poisson process is $N_i \sum_{j=1}^{N_L} m_j X_i$ $(y_i - y_j)$, where $N$ is the total population size and $m_j$ is the rate characterizing mutation $j \rightarrow i$ (we assume only one-point mutations: $m_j = m_0 = m$ if $d_{ij} = 1$, otherwise $m_j = 0$).

The dimensionality of the model, $C = N_k N_L^N_L$, is the same as in the deterministic case we studied.

We can also add transient strain-transcending immunity to the model, as hypothesized in Ferguson et al. (2003) and described in Minayev & Ferguson (2008). We assume that a host becomes temporarily immune to all strains upon infection but then loses this immunity at a rate $\gamma$. Dynamics of the temporarily immune populations required to model this is described by the equations

$$\frac{dr_i^{(k)}}{dt} = s_i^{(k)} \sum_{j \neq i, d_{ij} \leq k} \lambda_j - (\gamma + \mu) r_i^{(k)},$$  \hspace{1cm} (2.5)

Equations (2.1) and (2.4) retain their form and (2.2) is modified to include the term describing population returning from the temporarily immune to the susceptible state

$$\frac{ds_i^{(k)}}{dt} = \mu (1 - s_i^{(k)}) - s_i^{(k)} \sum_{j \neq i, d_{ij}} \lambda_j + \gamma r_i^{(k)}.$$  \hspace{1cm} (2.6)

The dimensionality of this extended model, $C = 2N_k N_L^N_L$, is approximately twice that of the model with no transient immunity.

3. SYSTEM DYNAMICS

Simulations were undertaken for two scenarios of pathogen genome complexity, one assuming seven loci and three alleles allowing for $3^7 = 2187$ strains in total, and the other assuming six loci and three alleles allowing for $3^6 = 729$ strains.

We continue to assume that the cross-immunity function $\gamma(d)$ has the form given in equation (2.3). We focus on examining dynamics for large $a$ ($a = 0.2 \ldots 0.5$) corresponding to three or six allele substitutions at different loci being required to completely escape prior immunity. Cross-immunity generated by one strain against strains only one allele different from that strain is very high ($b = 0.95$ or 0.9). These parameters are
chosen to be ‘influenza-like’, in that they might be expected to maximize diversifying selection. The birth/death rate and mutation rate of the virus are the same as for the deterministic model ($m = 0.014$ per year and $m = 10^{-3}$ per year). For all simulation runs, we assume that initially there is only one strain and after a period of approximately five months introduce another strain, which contains only two novel alleles. We equilibrate the model for 100 years (chosen by observing the period required for model equilibration) before analysing 200 years of model output.

The dynamics of the system is essentially different from the one of the deterministic model we studied before. In particular, regular periodic dynamics and equilibria are superseded by epidemic oscillations with widely varying amplitudes. An example time series for the proportion of infected population is portrayed in figure 1a, b. The average prevalence of infection grows as cross-immunity becomes more specific (i.e. for increasing $a$).

Mean strain lifetimes for the fully deterministic model are effectively infinite since only an intrinsic fitness deficit (lower $R_0$) can cause extinction. Allowing for stochastic extinction and mutation processes drastically alters this picture, leading to mean strain lifetimes of a few years. Mean strain lifetimes for a genotype with six loci and three alleles and $b = 0.9$ are plotted in figure 3a, b. Colour dots in these graphs are shown (population size: blue dots, 1 000 000 000; red dots, 300 000 000; green dots, 90 000 000). The maximum value of cross-immune response, $b$, is fixed and equal to 0.95. Mean strain lifetimes for a genotype with six loci and three alleles and $b = 0.9$ are plotted in figure 3a, b. Colour dots in these graphs

Figure 1. (a) Example time series for the proportion of infected in 1 billion population (virus genotype consists of six loci and three alleles). Curves of different colour show the dynamics of individual virus strains. (b) Temporal dynamics of the proportion of infected with all prevalent strains. The following parameter values were used: $\mu = 0.5$, $b = 0.9$, $R_0 = 2$, $\sigma = 50$.

Figure 2. Average strain lifetime as a function of the cross-immunity parameter $a$ for a genotype with seven loci and three alleles. $R_0$ is taken to be $(a,b) 2$, $(c,d) 4$ and $(e) 6$. Maximum cross-immunity value, $b$, is 0.95 in all the cases. Results for a recovery rate of $(a,b) \sigma = 50$ per year and $(c-e)$ for $\sigma = 100$ per year are shown (population size: blue dots, 1 000 000 000; red dots, 300 000 000; green dots, 90 000 000).
correspond to the results of single model runs and solid lines indicate general trends depending on the value of the cross-immunity parameter $a$. Owing to increasing prevalence and growing competition between strains, average strain lifetimes shorten as the cross-immune response becomes more specific (i.e. for increasing $a$). At large maximum values of cross-immune response ($b$ close to 0.95), this dependence is fairly weak (figure 2a–c) but becomes more distinct at $b \leq 0.9$ (figure 3a,b). Strain lifetimes also shorten (although only slowly) with decreasing population size (note that the extinction threshold for a strain is the prevalence of infection dropping below $1/N$, where $N$ is host population size). As expected, the lifetime of strains scales linearly with the generation time $1/\mu$, although this is only true where $\sigma \gg \mu$. Increasing $R_0$ at fixed $\sigma$ leads to a higher competition between strains thus also shortening their lifetimes. The dependence of strain lifetimes on $R_0$ appears to be almost linear within the studied parameter range.

More insight can be obtained from examining the distribution of strain lifetimes. The cumulative distribution function $F_N(t)$ is defined as the percentage of strains having lifetimes less than or equal to $t$ and density function $f_N(t)$ is the percentage of strains whose lifetimes fall within the interval $(t,t+\delta t)$, where $\delta t$ is a constant, which we assume to be equal to 1 year. For examples of $F_N(t)$ and $f_N(t)$ plotted for a population of 300 million, see figure 10a,b. Maximum strain lifetimes can be rather long: approximately 100 years when $R_0 = 2$ and $\sigma = 50$ per year, halving when the generation time is halved ($\sigma = 100$ per year). However, the proportion of strains surviving so long is very low. The vast majority of strains have lifetimes of a few years—when $\sigma = 50$ per year, approximately 84 per cent of strains survive less than 5 years for a population of 1 billion, rising to 88 per cent for a population of 90 million. For $\sigma = 100$ per year, the percentage going extinct in under 5 years rises to 95 per cent even in populations of 1 billion.

Since the genotype structure used in the model allows only for a limited number of strains, their stochastic re-emergence is inevitable. We examined how the average time of strain re-emergence depends on the number of alleles assumed per locus for a four loci antigenic ‘genome’. Figure 4 shows a weak linear trend when the number of alleles is varied between 3 and 8 (81–6561 different strains).

To characterize the overall diversity of the pathogen population through time, we use a frequency-weighted measure of diversity, $n_w(t) = \sum_k y_k(t)/y(t)^{\text{max}}$ where $y_k(t)$ is the proportion of infected with strain $k$ at time $t$, and $y(t)^{\text{max}}$ is the maximum of the set $\{y_k(t)\}$. If all strains have equal prevalence, this expression just reduces to the number of strains. Simulation shows that $n_w$ depends weakly on population size and the cross-immunity parameter $a$, but is more strongly affected by the recovery rate $\sigma$ and genome complexity (figure 5a,b). Varying $R_0$ with fixed $\sigma$ does not have a significant effect on $n_w$ in the range $R_0 \in [2,6]$. However, there is a strong negative correlation between diversity and generation time, as determined by $\sigma$ (figure 6a,b).

We now consider the impact of transient strain-transcending immunity on these dynamics. Figure 7 shows temporal dynamics of strain-specific infection prevalence for a pathogen with six antigenic loci, each with three alleles (i.e. 729 strains in total). Results are shown for mean duration of strain-transcending immunity of 0.5 year. The series of peaks plotted in different colours in bold demonstrates sequential strain replacement occurring on a short time scale (1–3 years), a phenomenon typical for influenza virus. On a longer time scale (more than 50 years), however, re-emergence of strains is inevitable due to the limited total number of strains being modelled.
Transient non-specific immunity also leads to shortening average strain lifetimes (figure 8) and lower diversity (figure 9). For the model with six-month long transient immunity, the reduction in strain lifetimes and diversity is 1.5- to twofold, compared with the model with the same base parameters but without non-specific immune response. Examining the distribution of strain lifetimes, figure 10, shows examples of $F_s(t)$ and $f_s(t)$ plotted for a population of 300 million for the models both with (red curve) and without (blue curve) transient immunity. Maximum strain lifetimes vary from 20 to 45 years depending on the duration of temporary immunity (the longer the duration the shorter is the maximum strain lifetime) and are generally 1.5- to 2.5-fold shorter than in the absence of transient immunity. The number of strains surviving so long is very few and all have very distinctive and similar dynamics: after an initial epidemic peak, a deep prevalence trough occurs that can last approximately one decade. During this trough, strains do not go completely extinct but, infecting a very small proportion of population ($10^{-6}$ to $10^{-5}$), survive to give rise to a new epidemic peak or series of peaks once population immunity to strain wanes. It is arguable that improving model realism via the inclusion of spatial heterogeneity and seasonality of transmission into the model would eliminate such long-lived strains, but in order to maintain model simplicity we do not explore such modifications here.

Phylogenetic trees generated from 50 years of model output also indicate that adding transient non-specific immunity gives more ‘flu-like’ evolutionary dynamics (figure 11), with a pattern of sequential replacement of fairly short-lived lineages seen (compared with much greater standing diversity for the model without transient immunity). These dynamics can also be seen by examining the degree of genetic drift over time.

Figure 5. Prevalence-weighted numbers of strains circulating yearly as a function of the cross-immunity parameter $\alpha$, for $R_0=2$, \( b=0.95 \) and \( b=0.9 \). Colour triangles correspond to the results for a recovery rate of $\sigma=50$ per year and circles to the results for $\sigma=100$ per year. (a) Results for a seven-loci, three-allele genotype and (b) for a six-loci, three-allele genotype are shown (population size: blue, 1 000 000 000; red, 300 000 000; green, 90 000 000).

Figure 6. Time dependence of the prevalence-weighted number of co-circulating strains (blue curve) and strain extinction rate (red curve). Both graphs are for 300 million population, $R_0=2$, $b=0.9$ and a genotype of six loci and three alleles. Recovery rates of (a) $\sigma=50$ per year and (b) $\sigma=100$ per year.

Transient non-specific immunity also leads to shortening average strain lifetimes (figure 8) and lower diversity (figure 9). For the model with six-month long transient immunity, the reduction in strain lifetimes and diversity is 1.5- to twofold, compared with the model with the same base parameters but without non-specific immune response. Examining the distribution of strain lifetimes, figure 10, shows examples of $F_s(t)$ and $f_s(t)$ plotted for a population of 300 million for the models both with (red curve) and without (blue curve) transient immunity. Maximum strain lifetimes vary from 20 to 45 years depending on the duration of temporary immunity (the longer the duration the shorter is the maximum strain lifetime) and are generally 1.5- to 2.5-fold shorter than in the absence of transient immunity. The number of strains surviving so long is very few and all have very distinctive and similar dynamics: after an initial epidemic peak, a deep prevalence trough occurs that can last approximately one decade. During this trough, strains do not go completely extinct but, infecting a very small proportion of population ($10^{-6}$ to $10^{-5}$), survive to give rise to a new epidemic peak or series of peaks once population immunity to strain wanes. It is arguable that improving model realism via the inclusion of spatial heterogeneity and seasonality of transmission into the model would eliminate such long-lived strains, but in order to maintain model simplicity we do not explore such modifications here.

Figure 7. Sample time series for the proportion of infected population (300 million population model, $R_0=2$, $\sigma=50$ per year; cross-immune response function, $\gamma(d)$, is defined by the parameters $a=0.45$ and $b=0.9$). A few peaks corresponding to different strains are plotted in bold showing sequential strain replacement. The duration of strain-transcending immunity is 0.5 year.
Figure 12a,b shows the number of genetic changes accumulated by strains over time for a number of runs of the models without and with strain-transcending immunity. In the absence of non-specific immunity, strain diversification occurs quite rapidly and in just 3–4 years the genetic drift curve reaches a plateau (caused by the finite numbers of strains in the model), whereas, for the model with non-specific immunity, the drift curve is more linear over time, similar to the one observed for influenza A. Eventually, even in that case the curve saturates due to the limited number of strains possible in the model. This is a common drawback for all models with low genotype complexity (e.g. Girvan et al. (2002) where a 32 bit string model was studied).

In summary, including transient non-specific immunity results in a model that better matches the observed epidemiological and evolutionary dynamics of human influenza A viruses, with realistically low levels of strain diversity and reasonable strain lifetimes, together with sequential strain replacement. However, even the semi-stochastic model without transient immunity is a closer match to influenza dynamics than the purely deterministic formulation of the same model (Minayev & Ferguson 2008).

4. CONCLUSIONS

This paper has introduced a new class of multi-strain disease models that retain the relatively parsimonious compartmental structure of previous models (Andreasen et al. 1997; Gupta et al. 1998; Minayev & Ferguson 2008) but accounts for finite population sizes by explicitly modelling strain extinction, and captures the stochasticity of the mutation process. The dynamics of these ‘semi-stochastic’ models are strikingly different from those of the entirely deterministic model, with large amplitude limit cycles and chaotic behaviour typically being replaced by oscillatory dynamics with a limited number of dominant strains circulating at any one time. The maximum lifetime of a strain can be long—with the 99th percentile lifespan being approximately 40 years (figure 10a, blue curves)—but the majority of strains go extinct within 5 years. Transient non-specific immunity shortens strain lifespan still further, reducing the upper 99th percentile lifetime to approximately 10 years (figure 10a, red curves).

Overall, we assessed the influence of population size, generation time, genome complexity and cross-immunity on system dynamics. As expected, average strain lifetime and frequency-weighted diversity depend approximately logarithmically on host population size, with both variables being more strongly affected by pathogen generation time (diversity being inversely proportional to generation time) and the number of antigenic loci and alleles assumed. Our analyses demonstrate that the processes underlying inter-strain competition and the complex epidemic dynamics resulting from such competition are strongly shaped by the processes that generate and
remove strains—namely mutation and extinction. This is especially true for diseases with short generation times and high mutation rates, such as influenza. Indeed, our results cast doubt on the hypothesis proposed in Recker et al. (2007) to explain the antigenic evolution of influenza A. This is because the diversity of antigenic types circulating at any time in the model proposed in that paper was restricted assuming a very small number of genetic loci determined influenza antigenicity and that no mutation occurred. Increasing the number of antigenic loci and alleles to more reasonable levels and incorporating strain extinction and stochastic mutation give very different dynamics, as illustrated in this paper. Strain recycling then only occurs by strains being regenerated via mutation (due to the finite size of the modelled antigenic genome)—a process that has never been observed in reality. Indeed, recent data from haemagglutinin inhibition tests comparing strains spanning the entire evolutionary history of H3N2 find no evidence of antigen recycling (Smith et al. 2004).

Our results demonstrate that incorporation of transient strain-transcending immunity reduces diversity and strain lifetime giving dynamics that more closely resembles those of human influenza A (e.g. sequential replacement of dominant strains) than models that do not include such immunity—in agreement with prior simulation studies (Ferguson et al. 2003). However, the models explored in this paper do not incorporate spatial heterogeneity or seasonality in transmission, both factors that have a substantial effect on strain establishment, competition and extinction (Rambaut et al. 2008; Russell et al. 2008). This limits the extent to which results can be quantitatively compared with epidemiological or evolutionary data for human influenza.

Figure 11. Comparison of example phylogenetic trees constructed from 50 years of output from the models (a) without transient non-specific immunity and (b) with transient immunity. Duration of transient immunity is 0.5 year and other model parameters are the same as in figure 9. Colour segments indicate times when strains are first found in circulation.

Figure 12. Graphs showing genetic drift for a model (a) without transient non-specific immunity and (b) with transient immunity plotted for a total of nine runs. For both models, 10-loci, two-allele genotype is considered and the other parameters are the same as in figure 9.
The semi-stochastic model we have developed here has the advantage of having relatively low computational complexity compared with full individually based simulations (Ferguson et al. 2003; Koelle et al. 2006), which makes rigorous exploration of parameter space more feasible. The key weakness of the model structure is that it allows only rather limited antigenic diversity to be considered as model dimensionality grows exponentially with the number of antigenic loci and as a power function of the number of alleles assumed. For that reason, extinct strains can randomly re-emerge in short time (25–50 years for the genotype sizes explored here), which somewhat limits the realism with which this type of model (whether stochastic or deterministic) can describe influenza evolution. The same shortcoming also applies to other compartmental frameworks, such as status-based models (Gog & Grenfell 2002). However, despite this limitation, the models developed here are potentially useful as descriptions of the diversity and evolution of other pathogens, most notably bacteria such as Neisseria meningitidis (Gupta et al. 1996) and viruses such as dengue (Ferguson et al. 1999). Extending the current framework to include recombination might also allow relatively parsimonious description of the population-level evolution of macroparasite infections, such as malaria.

REFERENCES

Andreasen, V., Lin, J. & Levin, S. A. 1997 The dynamics of cocirculating influenza strains conferring partial cross-immunity. J. Math. Biol. 35, 825–842. (doi:10.1007/s002850050079)

Bush, R. M., Bender, C. A., Subbarao, K., Cox, N. J. & Fitch, W. M. 1999 Predicting the evolution of human influenza A. Science 286, 1921–1925. (doi:10.1126/science.286.5446.1921)

Ferguson, N., Anderson, R. & Gupta, S. 1999 The effect of antibody-dependent enhancement on the transmission dynamics and persistence of multiple-strain pathogens. Proc. Natl Acad. Sci. USA 96, 790–794. (doi:10.1073/pnas.96.2.790)

Ferguson, N., Galvani, A. & Bush, R. 2003 Ecological and immunological determinants of influenza evolution. Nature 422, 428–433. (doi:10.1038/nature01509)

Girvan, M., Callaway, D. S., Newman, M. E. J. & Strogatz, S. H. 2002 A simple model of epidemics with pathogen mutation. Phys. Rev. E 65, 031915–1–031915–9. (doi:10.1103/PhysRevE.65.031915)

Gog, J. & Grenfell, B. 2002 Dynamics and selection of many-strain pathogens. Proc. Natl Acad. Sci. USA 99, 17 209–17 214. (doi:10.1073/pnas.252512799)

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. Nat. Med. 2, 437–442. (doi:10.1038/nm0496-437)

Gupta, S., Ferguson, N. M. & Anderson, R. M. 1998 Chaos, persistence and evolution of strain structure in antigenically diverse infectious agents. Science 280, 912–915. (doi:10.1126/science.280.5365.912)

Koelle, K., Cobey, S., Grenfell, B. T. & Pascual, M. 2006 Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. Science 314, 1898–1903. (doi:10.1126/science.1132745)

Minayev, P. & Ferguson, N. 2008 Improving the realism of deterministic multi-strain models: implications for modeling influenza A. Interface (doi:10.1098/rsif.2008.0333)

Rambaut, A., Pybus, O. G., Nelson, M. L., Viboud, C., Taubenberger, J. K. & Holmes, E. C. 2008 The genomics and epidemiological dynamics of human influenza A virus. Nature 453, 615–619. (doi:10.1038/nature06945)

Recker, M., Pybus, O. G., Nee, S. & Gupta, S. 2007 The generation of influenza outbreaks by a network of host immune responses against a limited set of antigenic types. Proc. Natl Acad. Sci. USA 104, 7711–7716. (doi:10.1073/pnas.0702154104)

Russell, C. A. et al. 2008 The global circulation of seasonal influenza A (H3N2) viruses. Science 320, 340–346. (doi:10.1126/science.1154137)

Smith, D. J., Lapedes, A. S., de Jong, J. C., Bestebroer, T. M., Rimmelzwaan, G. F., Osterhaus, A. D. M. E. & Fouchier, R. A. M. 2004 Mapping the antigenic and genetic evolution of influenza virus. Science 305, 371–376. (doi:10.1126/science.1097211)