Available online at www.sciencedirect.com
SCIENCE DIRECT®
Mathematical Biosciences 188 (2004) 17–28
Mathematical Biosciences
www.elsevier.com/locate/mbs

# Evolution and persistence of influenza A and other diseases

Simon A. Levin *, Jonathan Dushoff, Joshua B. Plotkin

*Department of Ecology and Evolutionary Biology, Princeton University, 203 Eno Hall, Princeton, NJ 08544 1003, USA*

Received 17 December 2002; received in revised form 13 August 2003; accepted 13 August 2003

**Abstract**

The evolution of the etiological agents of disease presents one of the greatest challenges for their control, and makes essential complementing standard epidemiological investigations with broader approaches that allow for evolutionary change. Given the stunning genetic diversity that is possible for many such agents, such as the influenza virus, it is impossible to represent all of the diversity manifest at the level of amino acid sequences. We show that drift-variant influenza strains naturally cluster into groups which are associated with functionally important epitopic regions. Dominant clusters typically replace each other every 2–5 years, and this feature is fundamental to the development of vaccination strategies. We furthermore show that stochastic fluctuations can greatly magnify small interference effects among strains, or even among subtypes, leading for example to competitive exclusion in situations where such effects would be unexpected based on the usual deterministic models. We suggest that this effect might be involved in the explanations of some persistent empirical anomalies.


*Keywords:* Disease evolution; Influenza; Hemagglutinin; Stochastic models

## 1. Introduction

The classical theory of epidemics is one of the jewels of mathematical biology, elegant in its form and providing insights and guidance into the management of infectious diseases. The basic model (Fig. 1) divides the host population into classes – typically susceptible (S), infectious (I) and removed (recovered and immune, R) – according to their disease status, and develops a set of differential equations to track changes in abundance (see [2]).

* Corresponding author. Tel.: +1-609 258 6880; fax: +1-609 258 6819.
*E-mail address:* simon@eno.princeton.edu (S.A. Levin).

0025-5564/$ - see front matter 
doi:10.1016/j.mbs.2003.08.010

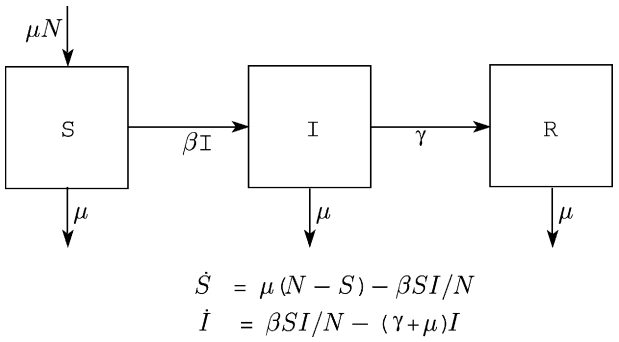


Fig. 1. A diagram of the S-I-R 'box' model. Individuals become infected (move from S to I) at a rate proportional to the number of infectious individuals in the population, the only non-linear process in the model. Additionally new individuals are born into the susceptible class, and all individuals die at rate $\mu$. The basic reproductive number for this model is given by $R_0 = \beta/(\mu + \gamma)$. Since the population size is constant, the equation for $R$ is redundant, and is omitted.

In the classical S-I-R model, with constant population size, no sustained oscillations are possible. Depending on the parameters and population size, the disease either dies out, or is sustained at an endemic level. If the duration of immunity is long compared to the duration of infectiousness (as in most human diseases) disease prevalence shows damped oscillations. But real diseases show persistent fluctuations, spatially and temporally, on a wide variety of scales. What then is lacking from the basic models, and how can they be adjusted to account for the persistence of fluctuations?

First of all, it is important to separate the various phenomena – fluctuations on the scale of days or weeks have very different explanations from those on the scale of years or generations. For influenza A, for example, there are clear tendencies for seasonal outbreaks in the Fall, although the actual date of emergence and spatial locus is somewhat unpredictable within the envelope of possibilities. These patterns apply both in the Northern Hemisphere and the Southern, shifted by approximately 6 months. The extent to which these patterns relate to the intrinsic properties of the virus in changing climatic conditions, to the social dynamics of children entering crowded schoolrooms, or to other factors remains a point of debate, however, as does the extent to which such exogenous factors are amplified by the oscillatory tendency of the disease system. Furthermore, it is not clear whether these are self-sustaining patterns within each hemisphere, in which case a puzzle is to explain how and where the virus survives in between epidemics, or whether the two hemispheres constantly re-seed each other through migrants. The truth probably involves both mechanisms.

Over longer time scales, diseases may oscillate due to evolutionary changes, combined with changes in the epidemiological status of the host population. Indeed, the continual evolution of pathogens such as the influenza A virus are crucial to their survival, involving an existential game of cat and mouse with their hosts. Understanding the role of evolution can thus be a powerful tool in controlling disease. For example, major new subtypes of the influenza A virus, typically caused by reassortment between an avian influenza virus and a human strain in a coinfected host, cause devastating pandemics on generational time scales. The 1918 Spanish Flu, for example, caused perhaps 40 million deaths [40] and great morbidity, largely among individuals in their twenties and thirties. The emergence of new subtypes through reassortment is termed 'shift', and raises a

large number of theoretical and practical challenges. Fluctuations also occur over shorter time scales, i.e. from year to year, due to mutation-based 'drift' variants. It is these changes that occupy most of the attention of those concerned with determining recommended vaccines on an annual basis. Indeed, despite the fact that the annual epidemics associated with drift variants are much less pronounced than when pandemics occur, their cumulative impact still greatly outweighs pandemics [37,38].

## 2. Oscillations in disease models

In general, there are a wide variety of mechanisms that can lead to oscillations in disease dynamics [31] These include:

- Stochastic factors,
- Seasonal forcing,
- Age structure, which can introduce implicit delays,
- Fluctuating population sizes, perhaps driven in part by disease dynamics,
- Non-linear incidence functions, in which the number of new cases does not increase in simple proportion to the number of susceptibles or number of infectives [27],
- Spatio-temporal interactions [5,20],
- Interactions among multiple diseases or strains.

Various attempts have been made to look at all of these mechanisms, and their influence upon fluctuations. Each has its own signature, and decomposing observed fluctuations into those pieces attributable to the various possible causes is a daunting challenge, made more difficult by the non-linear reinforcement of one mechanism by another. For example, Castillo-Chavez et al. [10] showed that introducing age structure alone, or two interacting strains of the influenza A virus alone, increased the tendency to oscillate but still resulted in damped oscillations. Coupling together multiple strains with age structure is thus analogous to coupling two damped oscillators, and was found to lead to sustained oscillations.

In more recent work, Andreasen et al. [3] found that four interacting strains, with appropriate patterns of cross-reactivity, could sustain oscillations without any other model complexities, such as age structure; and Lin et al. [26] showed that indeed this is possible even with as few as three strains, one of which interacts strongly with the other two, which in turn interact weakly with each other. Even with only three strains, the dynamical system is complicated, with eight levels of susceptibility (depending on what combination of strains an individual has experienced in the past) and 12 kinds of infected individuals (distinguished by current infection and past history). Thus, even for this simplified system, there are 20 coupled differential equations to be explored.

In all of this work, the definition of a strain is left vague. Mathematically, it refers to a genetic unit that can be treated as homogeneous in its properties, and has definable properties of cross-reactivity with other genetic units. If this unit is identified with subtypes, then the partitioning is in general too coarse – observed patterns of cross-reactivity among cocirculating subtypes appear to be weak to non-existent, although short-term immunological memory (or even the temporary isolation of infected individuals) can have surprisingly strong effects upon macroscopic dynamics

(see Section 4). On the other hand, if each unique RNA sequence is treated as a unique strain, one has both a hopelessly complicated problem, with literally thousands of equations [3,18,19], and furthermore with more detail than is appropriate to explain observed patterns. Thus, a relevant question is whether there is an intermediate scale, involving clusters of types (or 'quasispecies') that define the natural scale for explaining fluctuations on the order of one or several years. We turn to this question in the next section.

## 3. Genotype clusters in the influenza A virus

Theoretical work on the epidemiological dynamics of interacting disease types makes clear that it is important to clarify what exactly a functional type is. As pointed out in the last section, too much detail can create a hopeless morass of equations, while too little detail can miss crucial interactions. In this section, we turn our attention to determining an appropriate scale of strain aggregation for dealing with the annual epidemics of influenza A, and with changes that occur over a few seasons. In the next section, we look at longer time scales.

Within a given subtype, influenza A evolves at an astonishing rate, producing a large amount of RNA sequence diversity. Considerable evidence [13,14,34] suggests that rapidly evolving RNA viruses effectively experience selection as clusters rather than as individual genotypes. These clusters qualify as 'quasispecies' in the broad sense of a cluster of genotypes that may be seen as a unit of selection [14], and not in the narrower sense (often used in the HIV literature) of a cluster of genotypes within a single infected individual.

The drift variants of influenza A are produced by gradual substitutions to its genome, especially in its hemagglutinin gene. Hemagglutinin, or HA, is the primary target of the antigens produced by the immune system to neutralize an influenza infection. Recovery from influenza infection brings lasting immunity to the infecting genotype, but most people are susceptible to re-infection by a new drift variant within a few years. Thus antigenic drift requires that vaccines be updated annually to correspond to the dominant epidemic genotype(s) of HA.

The phenomenon of acquired immunity suggests that the influenza virus may be under frequency-dependent selection. Novel variants of influenza may initially enjoy a selective advantage due to the immunologically naive pool of human hosts; but after a given variant has caused an epidemic, it is then handicapped by the immunity it has stimulated in the human population. Phylogenetic reconstructions of HA evolution [16,17] confirm that modifications to the immunogenic parts of HA accrue at a dramatic rate. Those sites of HA involved in antigen determination also exhibit significantly more non-synonymous nucleotide substitutions than synonymous substitutions [8,21].

Although phylogenetic algorithms can estimate the evolutionary relationships among sequences, influenza phylogenies, even those using extensive databases, are often plagued by poor bootstrap values and instabilities of tree topology, which have been systematically studied by only a few authors [7,9,16]. The problem of resolving an evolutionary time series to the level of individual sequences is thus difficult, but perhaps unnecessary.

We have attempted to identify swarms of functionally related influenza genotypes by detecting a natural length scale at which HA amino acid sequences aggregate into clusters. We have studied

a database of 560 aligned HA1 sequences, each 987 nucleotides, from influenza viruses isolated between 1968 and 2000 [28,29,36].

In order to identify viral clusters we first use the Hamming metric to define the distance between two HA1 sequences. That is, we count the number of pairwise amino acids that differ between two aligned sequences. (Alternative metrics that weight the differences between amino acids according to their stereochemical properties yield similar results.)

For each threshold distance $d$, we define 'clusters' in the sequence database as the connected components of the graph $G_d$ that results from placing edges between sequences within a distance $d$ of each other. For each threshold distance $d$, we define the mean cluster size $C(d)$ as the first moment of the (discrete) distribution of connected component sizes of $G_d$, divided by the number of sequences. Equivalently, $C(d)$ equals the probability that two randomly chosen sequences lie in the same cluster. As we vary $d$ we obtain a hierarchy of graphs, from the completely disconnected graph to the fully connected graph.

We detect a natural, non-random scale of aggregration by inspecting the percolation behavior of the mean cluster size as $d$ varies. According to the fundamental theorem of continuum percolation [32], a large database of random sequences would exhibit a mean cluster-size curve with a single sharp phase transition from $C(d) = 0$ for small $d$ to $C(d) = 1$ for large $d$. Therefore, non-trivial plateaus in the cluster-size curve correspond to stable length scales at which the sequences form non-random clusters.

In order to produce the smooth cluster-size curve in Fig. 2, for each threshold distance, $\langle d \rangle$, we in fact average over many realizations of the clustering (cf. [35]). In each realization, we connect

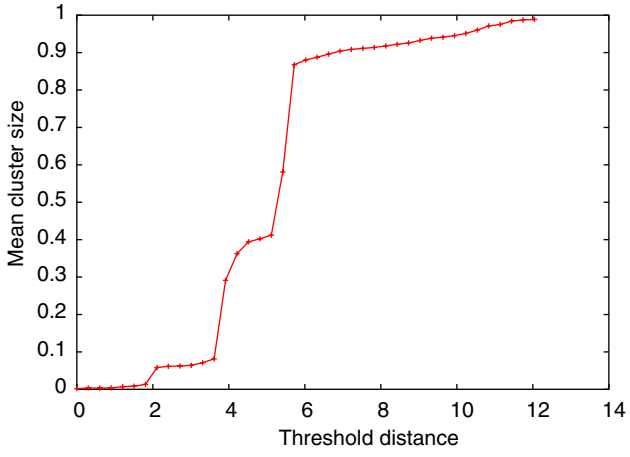


Fig. 2. The mean cluster-size curve for 560 sequences of HA1. This curve shows the relationship between the threshold distance $d$ (at which to connect two sequence into the same cluster) and the resulting normalized mean cluster size $C(d)$. Plateaus in the cluster-size curve correspond to stable length scales at which the sequences form non-random clusters. The smooth cluster-size curve results from averaging over 100 probabilistic Gaussian draws for each mean distance parameter $d$, with a 5% coefficient of variation (cf. [35]). The HA1 data exhibit two significant plateaus corresponding to clusterings at $d = 2$–3 and $d = 4$–5. The long tail for $d \geqslant 6$ corresponds to the gradual accumulation of outlier sequences. When $d = 2$, there are 174 resulting clusters with $C(2) = 0.0614$; at this scale, the mean cluster size is 34.4 sequences.

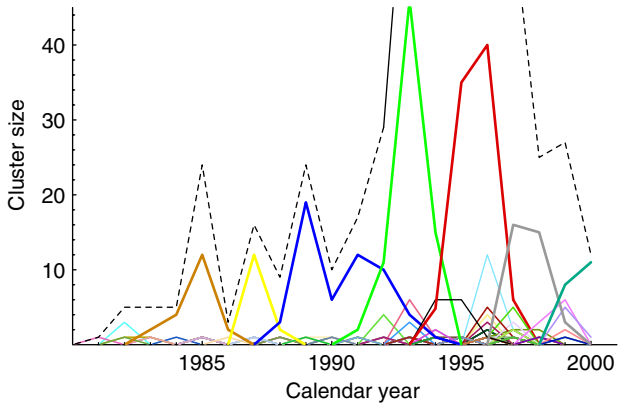


Fig. 3. The number of HA1 sequences within each cluster plotted as a function of calendar year of isolation. The clustering shown here corresponds to $d = 2$ amino acids (see Fig. 1). Each cluster is indicated by a different color, with the eight largest clusters shown in bold. The dashed line indicates the total number of isolates in the data set each year. The dominant sequence clusters tend to replace each other every 2–5 years. The dominant cluster in each year accounts for more than 25% of the sequences isolated that year. (The number of sequences each year does not reflect the severity of infections, but rather temporal biases in the sequence data set.)

two sequences by an edge if they lie within a randomly chosen distance $d$, with mean $\langle d \rangle$ and 5% coefficient of variation. (The connection distance $d$ is chosen from a truncated normal distribution.) Thus the cluster-size curve in Fig. 2 is, in essence, the average over many slight perturbations of the deterministic clustering method described above.

Fig. 2 shows that there is a natural, non-random partitioning of the HA sequences into disjoint clusters at a threshold distance of $d = 2$ amino acid changes. In other words, $d = 2$ provides the finest natural scale at which the sequences aggregate. (There is another non-random partition corresponding to $d = 4$, at which most sequences fall into three large clusters.) We interpret the existence of a natural scale of non-random aggregation among HA amino acid sequences as evidence that the influenza A virus forms viral quasispecies.

The time series of influenza sequence clusters strengthens our interpretation of viral clusters as selective units. Each of the 560 HA sequences is associated with a calendar year of viral isolation from an infected human. In Fig. 3 we plot the number of sequences in each cluster as a function of isolation year. Even though we did not use any information about isolation year when clustering the data, Fig. 3 shows that the resulting viral clusters are localized in time. There is no cluster that has members spanning more than seven collection years. Instead, dominant clusters of viral sequences tend to replace one another every 2–5 years, in agreement with the time scale of dominant antigenic replacements [11].

Note that the cluster time series seen in Fig. 2 is not periodic within the time-span of two decades. Once HA evolves away from a given region of sequence space, it does not later revisit that region. This result, seen here in terms of cluster structure, is consistent with the one-trunk phylogenetic reconstructions of HA [16].

Fig. 2 demonstrates that the general pattern of influenza A drift evolution is characterized by non-random clusters of highly related genotypes that replace each other every 2–5 years. The

notion of an influenza 'strain', which has hitherto been left unspecified, may arguably now be identified with the observed dominant, non-random sequence clusters. At this scale of precision, we see that influenza A virus exhibits significant evolutionary change on a 2–5 year time scale, as opposed to the annual periodicity of epidemics.

Our analysis has identified viral clusters on the basis of the Hamming distance between amino acid sequences. The Hamming metric on sequences has been used as an effective model of antigenic distances or B-cells in general [23] and for influenza in particular [39]. Nevertheless, comparison of sequence data to direct immunological assays should be used to quantify the exact correlation between amino acid composition and antigenic properties [24]. In other words, we have good reason to believe that all the sequences within a given cluster have similar patterns of cross-reactivity with outside sequences; but direct immunological confirmation is still required.

Once a natural scale of aggregation has been identified, the resulting empirical clusters can be inspected in greater detail. Previously, we have analyzed the spatio-temporal spread of sequences within a cluster, and we have discussed the potential utility of HA clustering for the optimal design of annual influenza vaccines [36]. We have also investigated patterns in the variation, cluster by cluster, of the dominant antibody-combining regions of HA, which may hold clues to the mechanism of influenza's antigenic plasticity. For the purposes of the present discussion, however, it suffices to summarize that (i) we can detect non-random scales of viral sequence aggregation; (ii) at the finest such scale, clusters are localized in time; and (iii) dominant clusters replace one another every 2–5 years (as seen in Fig. 3).

The work described in this section provides a natural way to cluster strains in order to explore the evolution and effects of patterns of cross-reactivity in the influenza A virus. The dominant clusters that have been identified can be associated with key epitopic regions, which determine how the virus attaches to host cells. Given that there are a small number of such regions, it is reasonable to expect them to wax and wane in importance, demonstrating recurrent behavior over decadal time scales.

## 4. Demographic stochasticity, seasonality and viral competition

As pointed out in Section 2, one of the factors that can reinforce the damped oscillations of the S-I-R model, and allow persistent fluctuations, is stochasticity. A natural form of stochasticity to consider is *demographic* stochasticity, or the stochasticity that arises purely from considering the population to be composed of discrete individuals: thus the instantaneous rate of new infections (from the determistic model) becomes an instantaneous probability that a single individual will be infected.

Demographic stochasticity has surprisingly large effects in S-I-R models, particularly when the duration of immunity is long compared to the duration of infectiousness. Fig. 4 shows time series for the number of infectious individuals in a deterministic S-I-R model, and in the corresponding model with demographic stochasticity. While the deterministic model shows damped oscillations, oscillations in the stochastic model persist. The period of oscillations in the stochastic model is predicted very well by the period in the deterministic model, while the long-term magnitude of oscillations (assuming the disease has not gone stochastically extinct) is predicted very well by

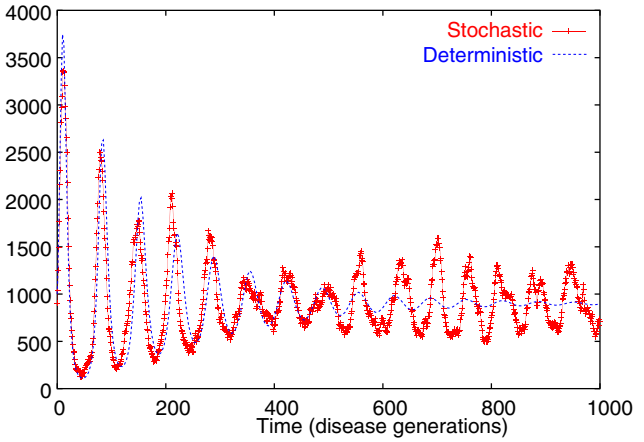


Fig. 4. The deterministic S-I-R model and one realization of the stochastic S-I-R model with the same parameters and initial conditions. Even though the equilibrium number of infectious individuals is on the order of thousands, demographic stochasticity leads to large, persistent fluctuations in the proportion infectious. The stochastic model is realized by converting each of the arrows in Fig. 1 into an instantaneous rate of birth, death, infection or recovery.

simple diffusion approximations [33]. Simple approximations fail, however, to capture the mean time to extinction of stochastic S-I-R systems.

Since influenza shows annual cycles in prevalence, it is natural to ask about seasonal forcing, and the extent to which exogenous forcing can amplify an endogenous tendency to oscillate. This question has been explored in tremendous detail for several childhood diseases, most notably measles [4,22]. There has been less work in this direction with influenza, both because there are fewer data available about influenza prevalence, and because the complication of drift evolution makes the question much more difficult.

The endogenous period of oscillation is critical to understanding how disease models are affected by seasonal forcing. For diseases with relatively high basic reproductive number $R_0$, and where the ratio of the duration of immunity to the duration of infection is high compared to $R_0$, the natural period of oscillation near the equilibrium is known to be approximately $2\pi\sqrt{ad}$, where $d$ is the length of the cycle of infection (latency plus infectiousness) and $a$ is the average age of first infection. The age at first infection can be further approximated as $a = 1/(R_0\omega)$, where $\omega$ is the rate at which people exit the recovered class (through death or loss of immunity). Taking values of 6–10 days for $d$ and 4–12 months for $a$ yields a natural period of between 0.5 and 1 year. The strong seasonal behavior observed in influenza is more consistent with values in the upper end of this range, which show more regular seasonal oscillations, and also stronger magnification of forcing (see Fig. 5).

Stochastic models also shed light on competitive interactions, including those between drift variants within a subtype of the influenza A virus, as discussed above, and between subtypes. At least in some cases, stochastic fluctuations can interact with other model factors to magnify greatly the effects of small competitive interactions.

Competing strains can interact if immune responses to one strain affect the ability of others to invade. Immune interaction can take the form of 'direct interference', whereby a current infection

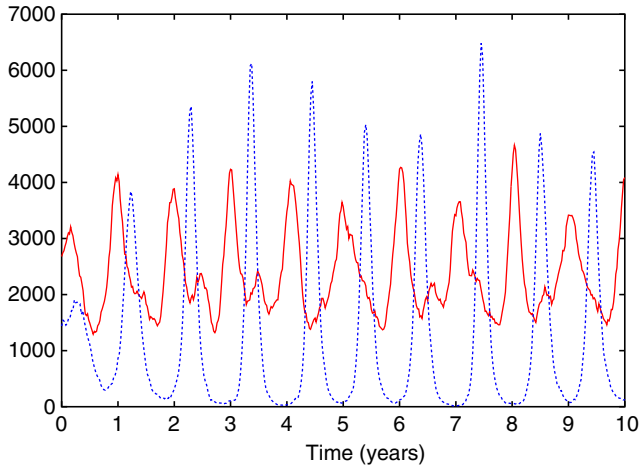


Fig. 5. Time series for the stochastic S-I-R model with annual forcing and different delays. In both simulations $N = 500000$ and $R_0$ varies sinusoidally between 9 and 11. The red curve has an infectious cycle of 0.02 year and duration of immunity of 4 years, giving a natural period of around 0.56 year; the blue curve has an infectious cycle of 0.025 year and duration of immunity of 8 years, giving a natural period of around 0.99 year.

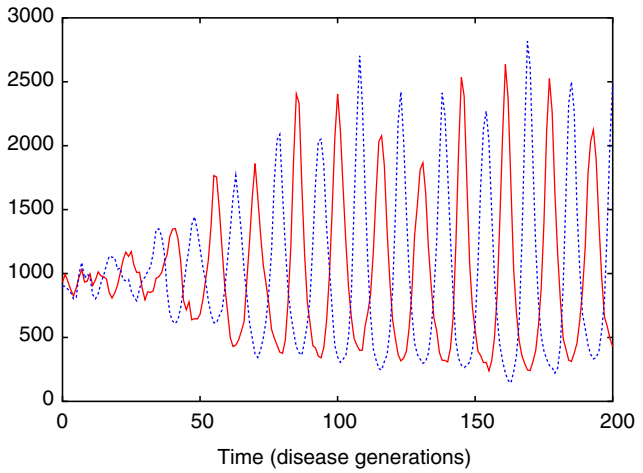


Fig. 6. Time series for a stochastic S-I-R with age structure, showing two strains that compete only by direct interference. Although the effect of these strains on each other is small at the individual level, it is very large at the population level.

with one strain precludes infection by other strains; or of 'cross-immunity', whereby the long-term immunity to a strain from which an individual has recovered reduces the ability of other strains to infect that individual.

Fig. 6 shows a stochastic simulation with two strains that have no cross-immunity, but only direct interference – so that an individual cannot be simultaneously infected by both strains. Since each strain infects less than 1% of the population at any given time, simple deterministic approximations would predict a negligible effect of each strain on the other. When stochasticity is

combined with age structure (in the form of normally distributed lifetimes), however, the two strains rapidly begin oscillating out of phase, greatly increasing the magnitude of oscillations, and driving one strain extinct in a relatively short time, despite the fact that either strain could persist for quite a long time on its own.

Thus, a very small competitive interaction at the individual level, which would have negligible effects in a deterministic approximation, has a very large effect at the population level when individuals are treated discretely. While more study is needed to determine the generality of this phenomenon, it clearly has potentially important implications to understanding flu 'shifts' and the riddle of why the old subtype is usually observed to go extinct when a new subtype comes in, despite the fact that the measured cross-immunity between different subtypes is very small.

## 5. Discussion

For many diseases, the key to persistence and the bane of control is their continual and rapid evolution. Diseases such as influenza A, which confers lifelong immunity to survivors of a particular strain, nonetheless reemerge year after year, causing tremendous morbidity and mortality, through an ability to change form and escape host immune systems. Understanding this evolutionary change is fundamental to management, and to reduction in the disease's toll. Influenza A is not unique in this – other diseases such as malaria, AIDS and hepatitis also can be caused by a variety of agents that are variants on a single theme.

There is, thus, a compelling need to extend the conventional modeling framework for infectious diseases to deal with evolutionary changes in the infectious agent. This is not uncharted territory, and has been an active area of research for 30 years or more [1,3,6,25,30]. Still, problems remain regarding how best to represent antigenic diversity. For example, in the case of the hemagglutinin gene of influenza A, thousands of different drift variants can be identified from infected individuals in as short a time as a decade. To model the dynamics of such an assemblage, each of which interacts uniquely with each other, is a task obviously beyond computation.

It is crucial, therefore, to find ways to partition variants into a small number of clusters, so that the variants within each cluster can be treated as functionally equivalent, interacting strongly among themselves and more weakly with strains in other clusters. Such techniques have long been used informally for a variety of diseases, for example myxomatosis [15], and indeed form the basis for vaccine choice [11,12] for influenza A. Still, as molecular techniques become more powerful, and computational algorithms more available, it is worthwhile to develop formal approaches to the problem of clustering strains, and to investigating the relationship between genetic sequences of hemagglutinin and their immunological properties.

We have shown that indeed it is possible to separate the near-infinity of strains into a small number of clusters, associated with the key epitopic regions of hemagglutinin (see also [36]). This work then provides a foundation for multitype modeling of infectious dynamics, and for guiding vaccine choice.

Ultimately, the difficulties faced in traditional epidemiological modeling relate to the necessity of lumping individuals into categories. With even 3 virus types, there are 20 categories of individuals; with 4, the number increases to 48. An alternative way to proceed is to develop individual-based models, in which unique life histories can be followed, and stochastic influences

easily incorporated. We show that such models can sustain oscillatory behavior when their deterministic equivalents cannot. Most importantly, these individual-based models illuminate the role of cross-reactivity among clusters, or among subtypes. Remarkably, it appears that only minimal interaction among strains or clusters can translate into strong influences on oscillatory behavior, and on patterns of cocirculation, at the population level. These results beg for deeper investigation, but may shed light on the puzzle of why subtypes, which interact only weakly, typically replace one another in temporal sequences. Similarly, we find that only minimal amounts of seasonal forcing can translate into large oscillations in disease prevalence.

## Acknowledgements

The authors are grateful for valuable discussions with Viggo Andreasen, David Earn, Ellis McKenzie, Tom Reichert, Lone Simonsen, and Martin Weigert, and for the comments of two anonymous reviewers. We thank Los Alamos National Laboratory and everybody who has contributed to the Influenza Sequence Database, including those who have contributed unpublished sequences. This work was supported by a grant from the National Institutes of Health, number 1-R01-GM60729-01 to S.A.L.; J.B.P. acknowledges support from the the National Science Foundation and the Burroughs Wellcome Fund award number 1001782.

## References

[1] R.M. Anderson, R.M. May, Population biology of infectious diseases 1, Nature 280 (1979) 361.
[2] R.M. Anderson, R.M. May, Infectious Diseases of Humans: Dynamics and Control, Oxford Science Publications, Oxford, 1991.
[3] V. Andreasen, J. Lin, S.A. Levin, The dynamics of cocirculating influenza strains conferring partial cross-immunity, J. Math. Biol. 35 (1997) 825.
[4] B. Bolker, B.T. Grenfell, Space, persistence and dynamics of measles epidemics, Philos. Trans. R. Soc. Lond. Ser. B—Biol. Sci. 348 (1995) 309.
[5] B.M. Bolker, B.T. Grenfell, Impact of vaccination on the spatial correlation and persistence of measles dynamics, Proc. Nat. Acad. Sci. USA 93 (1996) 12648.
[6] H.J. Bremermann, J. Pickering, A game-theoretical model of parasite virulence, J. Theor. Biol. 100 (1983) 411.
[7] R.M. Bush, Opinion: predicting adaptive evolution, Nat. Rev. Genet. 2 (2001) 387.
[8] R.M. Bush, W.M. Fitch, C.A. Bender, N.J. Cox, Positive selection on the H3 hemagglutinin gene of human influenza virus A, Mol. Biol. Evol. 16 (1999) 1457.
[9] R.M. Bush, C.B. Smith, N.J. Cox, W.M. Fitch, Effects of passage history and sampling bias on phylogenetic reconstruction of human influenza A evolution, Proc. Nat. Acad. Sci. USA 97 (2000) 6974.
[10] C. Castillo-Chavez, H.W. Hethcote, V. Andreasen, S.A. Levin, W.M. Liu, Epidemiological models with age structure, proportionate mixing, and cross-immunity, J. Math. Biol. 27 (1989) 233.
[11] N.J. Cox, C.A. Bender, The molecular epidemiology of influenza viruses, Semin. Virol. 6 (1995) 359.
[12] N.J. Cox, K. Subbarao, Influenza, Lancet 354 (1999) 1277.
[13] E. Domingo, J.J. Holland, C.K. Biebricher, Quasispecies and RNA Virus Evolution: Principles and Consequences, R.G. Landes Co., Austin, 2002.
[14] E. Domingo, L. Mededez-Arias, M. Quinones-Mateu, A. Holguin, M. Gutierrez-Rivas, M. Martinez, J. Quer, I. Novella, J. Holland, Viral quasispecies and the problem of vaccine-escape and drug-resistant mutants, Prog. Drug Res. 48 (1997) 99.

[15] F. Fenner, I.D. Marshall, Occurrence of attenuated strains of myxoma virus in Europe, Nature 176 (1955) 782.
[16] W.M. Fitch, R.M. Bush, C.A. Bender, N.J. Cox, Long term trends in the evolution of H(3) HA1 human influenza type A, Proc. Nat. Acad. Sci. USA 94 (1997) 7712.
[17] W.M. Fitch, R.M. Bush, C.A. Bender, K. Subbarao, N.J. Cox, Predicting the evolution of human influenza A, J. Hered. 91 (2000) 183.
[18] J.R. Gog, B.T. Grenfell, Dynamics and selection of many-strain pathogens, Proc. Nat. Acad. Sci. USA 99 (2002) 17209.
[19] J.R. Gog, J. Swinton, A status-based approach to multiple strain dynamics, J. Math. Biol. 44 (2002) 169.
[20] B.T. Grenfell, O.N. Bjornstad, J. Kappey, Travelling waves and spatial hierarchies in measles epidemics, Nature 414 (2001) 716.
[21] Y. Ina, T. Gojobori, Statistical analysis of nucleotide-sequences of the hemagglutinin gene of human influenza A viruses, Proc. Nat. Acad. Sci. USA 91 (1994) 8388.
[22] M.J. Keeling, B.T. Grenfell, Understanding the persistence of measles: Reconciling theory simulation and observation, Proc. R. Soc. Lond. Ser. B—Biol. Sci. 269 (2002) 335.
[23] V. Kohler, R. Puzone, P.E. Seiden, F. Celada, A systematic approach to vaccine complexity using an automaton model of the cellular and humoral immune system, Vaccine 19 (2000) 862.
[24] A. Lapedes, R. Farber, The geometry of shape space: application to influenza, J. Theor. Biol. 212 (2001) 57.
[25] S. Levin, D. Pimentel, Selection of intermediate rates of increase in parasite-host systems, Am. Natural. 117 (1981) 308.
[26] J. Lin, V. Andreasen, S.A. Levin, Dynamics of influenza A drift: the linear three-strain model, Math. Biosci. 162 (1999) 33.
[27] W.M. Liu, H.W. Hethcote, S.A. Levin, Dynamic behavior of epidemiologic models with nonlinear incidence rates, J. Math. Biol. 25 (1987) 359.
[28] Los Alamos National Labs. Influenza sequence database. Available from <http://www.flu.lanl.gov/>, World wide web site.
[29] C. Macken, H. Lu, J. Goodman, L. Boykin, The value of a database in surveillance and vaccine selection, in: A.D.M.E. Osterhaus, N. Cox, A.W. Hampson (Eds.), Options for the Control of Influenza IV, Elsevier, Amsterdam, 2001.
[30] R.M. May, R.M. Anderson, Population biology of infectious diseases 2, Nature 280 (1979) 455.
[31] R.M. May, R.M. Anderson, Periodicity in epidemiological models, in: S.A. Levin, T.G. Hallam, L.J. Gross (Eds.), Applied Mathematical Ecology, Springer, Berlin, 1989, p. 193.
[32] R. Meester, R. Roy, Continuum Percolation, Cambridge University, Cambridge, 1996.
[33] I. Nasell, On the time to extinction in recurrent epidemics, J. R. Stat. Soc. Ser. B—Stat. Methodol. 61 (1999) 309.
[34] M.A. Nowak, R. May, Virus Dynamics: The Mathematical Foundations of Immunology and Virology, Oxford University, Oxford, 2000.
[35] J.B. Plotkin, J.M. Chave, P.S. Ashton, Cluster analysis of spatial patterns in Malaysian tree species, Am. Natural. 160 (2002) 629.
[36] J.B. Plotkin, J. Dushoff, S.A. Levin, Hemagglutinin sequence clusters and the antigenic evolution of influenza A virus, Proc. Nat. Acad. Sci. USA 99 (2002) 6263.
[37] L. Simonsen, M.J. Clarke, L.B. Schonberger, N.H. Arden, N.J. Cox, K. Fukuda, Pandemic versus epidemic influenza mortality: a pattern of changing age distribution, J. Infect. Dis. 178 (1998) 53.
[38] L. Simonsen, K. Fukuda, L.B. Schonberger, N.J. Cox, The impact of influenza epidemics on hospitalizations, J. Infect. Dis. 181 (2000) 831.
[39] D.J. Smith, S. Forrest, D.H. Ackley, A.S. Perelson, Variable efficacy of repeated annual influenza vaccination, Proc. Nat. Acad. Sci. USA 96 (24) (1999) 14001.
[40] J.K. Taubenberger, A.H. Reid, T.G. Fanning, The 1918 influenza virus: a killer comes into view, Virology 274 (2000) 241.