Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Theoretical Biology 222 (2003) 437–445
Journal of Theoretical Biology
www.elsevier.com/locate/jtbi

Traveling waves in a model of influenza A drift

Juan Lin[a,*], Viggo Andreasen[b], Renato Casagrandi[c], Simon A. Levin[d]

[a] Department of Physics, Washington College, 3000 Washington Avenue, Chestertown, MD 21620, USA
[b] Department of Mathematics and Physics, Roskilde University, DK-4000 Roskilde, Denmark
[c] Dipartimento di Elettronica e Informazione, Politecnico di Milano, 20133 Milano, Italy
[d] Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA

Received 8 August 2002; accepted 31 December 2002

Abstract

Between major pandemics, the influenza A virus changes its antigenic properties by accumulating point mutations (drift) mainly in the RNA genes that code for the surface proteins hemagglutinin (HA) and neuraminidase (NA). The successful strain (variant) that will cause the next epidemic is selected from a reduced number of progenies that possess relatively high transmissibility and the ability to escape from the immune surveillance of the host. In this paper, we analyse a one-dimensional model of influenza A drift (Z. Angew. Math. Mech. 76 (2) (1996) 421) that generalizes the classical SIR model by including mutation as a diffusion process in a phenotype space of variants. The model exhibits traveling wave solutions with an asymptotic wave speed that matches well those obtained from numerical simulations. As exact solutions for these waves are not available, asymptotic estimates for the amplitudes of infected and recovered classes are provided through an exponential approximation based on the smallness of the diffusion constant. Through this approximation, we find simple scaling properties to several parameters of relevance to the epidemiology of the disease.

Keywords: Influenza A drift; Cross-immunity; Mutation; Traveling wave

1. Introduction

The human influenza A virus has been responsible for three major pandemic outbreaks in the last century. The pandemics are usually caused by the appearance of a new subtype (*antigenic shift*) having immunologically different genes coding for at least one of the surface proteins hemagglutinin (HA) and neuraminidase (NA) (Smith and Palese, 1989; Webster et al., 1992; Cox and Subbarao, 2000; Earn et al., 2002). Between pandemic outbreaks, the virus is able to produce annual epidemics of varying degrees of virulence by changing its amino acid composition in the active areas (epitopes) of the HA and NA proteins (Plotkin et al., 2002). This process, known as *antigenic drift*, can be examined at the genotype level by looking at the evolution of phylogenetic trees (Fitch et al., 1991, 1997; Bush et al., 1999). In the case of influenza A, the evolution proceeds along a main trunk with short side branches containing lineages that are not selected for. For the H3N2 subtype, the HA1 region of the HA protein is changing at the rate of nearly six nucleotide substitutions per genome per year, which is significantly higher than the silent rate of two nucleotides per genome per year in the non-structural gene (NS) (Buonagurio et al., 1986; Fitch et al., 1997). The rapid evolution of the HA1 region is the result of selective forces operating on the host (neutralizing antibodies and cell-mediated immune responses) and population levels (herd immunity). An epidemiological model must simplify the bewildering complexity of genotypic responses and map the evolution of variants on a phenotype space. There is no clear correlation among these two spaces, as linear ordering in the genome sequence does not reveal the full three-dimensional (3D) antigenic structure after protein folding. In building the model, we will assume that such phenotype space exists and that the virus drifts along a one-dimensional (1D) axis x of variants that mimics its evolution along the main trunk of the phylogenetic tree. Our model differs from Pease's (1987) 'evolutionary

*Corresponding author. Tel.: +1-410-778-7732; fax: +1-410-778-7275.
E-mail address: jlin2@washcoll.edu (J. Lin).

0022-5193/03/$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved.
doi:10.1016/S0022-5193(03)00056-0

epidemiology' and other versions (Girvan et al., 2002) in that while their models describe the effect of slow antigenic changes in a single variant on the dynamics of drift, they do not explain what causes drift. In our model, mutation is the driving force behind drift and this drift is constrained to move unidirectionally because of the structure of cross-immunity in the host population.

Instead of paying attention to the virus itself, we will concentrate on the number of infected hosts in the population (Levin and Pimentel, 1981). Several considerations will guide us in setting up the model. First, there is strong evidence that immunity of the host to reinfection by the same strain of influenza A is lifelong (Couch and Kasel, 1983). On a second challenge, memory B (long-term memory) and T (short-term memory) cells can be activated to reduce virus concentrations to levels that are harmless to the host. We incorporate this fact by allowing infected individuals to transmit the disease only to hosts that either have recovered from infections with variant types $y<x$ or are fully susceptible. Second, hosts recovered from infection with variant x exhibit a certain degree of cross-protection to challenges by related variants. This *cross-immunity* is manifested in HA inhibition assays (Levine, 1992; Both et al., 1983; Xu et al., 1993) and further corroborated by studies in closed populations (Potter et al., 1977; Larson et al., 1978). In the model, we introduce a kernel $K((x-y)/a)$ that reduces susceptibility to infection when y approaches x to within a distance a, while attaining almost full strength when the distance between x and y is significantly larger than a. An equivalent formulation ensues if instead of susceptibility we pay attention to transmissibility. Third, a full model must include the immunological history of the host (Andreasen et al., 1997). This can be done by dividing the population into classes, each carrying a signature characterizing current and all past infections of its members. In the model we analyse, these complications are avoided. We keep track of only the current and most recent past infection while adding a discrete fully susceptible class with a birth rate that exactly matches deaths from all classes. In Section 2, we outline an extension of the drift model first proposed by Andreasen et al. (1996) and estimate the asymptotic speed of the wave. In Section 3, we briefly discuss the threshold condition, $R_0>1$, necessary for the onset of an epidemic outbreak. The basic reproduction number R_0 (Anderson and May, 1991), is defined as the number of secondary infections caused by a single infected host in a totally susceptible population. We also establish conditions for the existence of traveling waves with amplitudes that vanish as $x\to\pm\infty$. In phase space, this means that the solution is a trajectory that begins and ends at the origin, or equivalently, a homoclinic orbit (Billingham and King, 2000). Furthermore, we explain why the numerical value of the wave speed as a function of the total population, may be below the asymptotic estimate. In Section 4, we develop an asymptotic approximation, based on the smallness of the diffusion constant, to obtain approximate solutions for the normalized densities of infected and recovered classes in the frame of reference moving at the speed of the wave front.

2. Drift model in phenotype space

The proposed model generalizes the classical SIR model (Anderson and May, 1991) by approximating drift as a diffusion process on a 1D axis of variant types. Let $S(t)$ be the number of fully susceptible hosts at time t, $i(x,t)$ the distribution of infected hosts, carrying variant x, and $r(x,t)$ the distribution of hosts who have recovered, and are immune to, variant x. We define the total population of infected and recovered as $I(t)=\int_{-\infty}^{\infty} i(x,t)\,dx$ and $R(t)=\int_{-\infty}^{\infty} r(x,t)\,dx$. The drift process is a local one and in a discrete space of variant types j it takes on the form

$$\tfrac{1}{2}[pi(j+1,t)+pi(j-1,t)]-pi(j,t),$$

where p is the mutation rate in antigenic space (Sasaki, 1994). Passing to the continuous limit, we find the equivalent expression

$$D\frac{\partial^2 i(x,t)}{\partial x^2},$$

where $D=p/2$. The per capita growth rate of infection by hosts carrying variant x is given as

$$bN\left[\int_{-\infty}^{x} K\left(\frac{x-y}{a}\right) r(y,t)\,dy+S(t)\right],$$

where the constant coefficients b and N are the contact rate and the total population. The susceptibility kernel K weighs the recruitment of recovered hosts to any $y<x$ by their closeness (in the antigenic sense) to x. The constant a is defined as the value of x such that susceptibility rises to half of its maximum value, $K(1)=\frac{1}{2}$. Host susceptibility is weakest within a distance of order a near x and rises monotonically with distance, e.g. $K=0$ at $y=x$ and $K\to 1$ when $|x-y|\gg a$. With the above expressions, the dynamics of drift and infection can be captured by the following system of equations:

$$\frac{dS(t)}{dt}=\mu(N-S(t))-bNS(t)I(t), \tag{1a}$$

$$\frac{\partial i(x,t)}{\partial t}=i(x,t)bN\int_{-\infty}^{x} K\left(\frac{x-y}{a}\right) r(y,t)\,dy-(\nu+\mu)i(x,t)+bNS(t)i(x,t)+D\frac{\partial^2 i(x,t)}{\partial x^2}, \tag{1b}$$

$$\frac{\partial r(x,t)}{\partial t} = -r(x,t)bN\int_x^\infty K\left(\frac{x-y}{a}\right)i(y,t)\,dy + vi(x,t) - \mu r(x,t). \quad (1c)$$

The constants $1/v$ and $1/\mu$ are the average infectious period and host lifetime, respectively. One can check directly that the total population N is constant by directly integrating the i and r equations over the whole x-axis and then adding up all three equations. The limits of integration in Eq. (1b) have been set so that recovered individuals at x can only be recruited by infected individuals carrying strains $y > x$. The asymmetric limits of integration prevent hosts from being continuously reinfected with the same strain. As the total population N remains constant we can rescale S, i and r as fractions of N, time in units of $1/(v+\mu)$ and distance along x in units of a. Therefore, let us define a new set of scaled variables,

$$S \to S/N, \quad i \to i/N, \quad r \to r/N, \quad t \to (v+\mu)t, \quad x \to x/a,$$

$$R_0 = bN/(v+\mu), \quad d = D/((v+\mu)a^2), \quad e = \mu/(v+\mu)$$

and rewrite system (1) as

$$\frac{dS(t)}{dt} = e(1-S(t)) - R_0 S(t) I(t), \quad (2a)$$

$$\frac{\partial i(x,t)}{\partial t} = R_0 i(x,t)\int_{-\infty}^x K(x-y)r(y,t)\,dy + R_0 S(t) i(x,t) - i(x,t) + d\frac{\partial^2 i(x,t)}{\partial x^2}, \quad (2b)$$

$$\frac{\partial r(x,t)}{\partial t} = -r(x,t)R_0\int_x^\infty K(x-y)i(y,t)\,dy + (1-e)i(x,t) - er(x,t). \quad (2c)$$

In Table 1, we provide some useful data about influenza A that will be used later on.

The averages of $v \approx 70$/year and $\mu \approx 1/(70\text{year})$ indicate the wide range of time-scales in influenza dynamics. The constant a can be estimated from Pease (1987), who reports that the probability of reinfection increases approximately linearly with time after the last infection, at least for an interval lasting a few years. This probability is about 30–40% after 5–6 years (Potter et al., 1977) and is proportional to the rise in susceptibility. Assuming annual epidemics, we estimate a to be of the order of 6–9 variants, which is the distance $|x-y| = a$ associated with a rise in susceptibility to 50% of the maximum. This parametrization of K determines the small diffusion constant d once we identify the speed of the wave with the speed of epidemic outbreaks, ≈ 1 variant per year. Further evidence for these values of a can be found in estimates of the ratio of infectious period over the average time T between infection and reinfection. This ratio is ≈ 0.002–0.004 (Pease, 1987) giving a value of $T \approx 7$–8 years or an infection in the host for every 7–8 variants sweeping the population. In the following, we will assume $d \ll 1$ in our calculations, and later verify that the assumption is consistent with the estimates for the asymptotic speed of the wave.

In the next section, we will establish under which conditions a steady traveling wave can be sustained. Here, we assume its existence and determine the speed by looking at the leading edge of the wave, where $i(x,t)$ is small. We first change variables to the co-moving frame, $u = x - ct$. In the infinite line of variants we will be looking for steady-state solutions that satisfy the boundary conditions, $(i(u), r(u)) \to (0,0)$ as $|u| \to \infty$. For large u, the integral term reaches the asymptotic value R (the total population of recovered hosts) and using a trial solution, $i \sim e^{-\lambda u}$, we derive the dispersion equation (Murray, 1989),

$$c = \frac{R_0(R+S) - 1 + d\lambda^2}{\lambda}.$$

As lambda is real and positive (other cases violate either $i(u) \geqslant 0$, or $i(u) \to 0$ as $u \to \infty$) we observe that the range of possible speeds is $c \geqslant c_0$, where c_0 is the minimum speed. If the initial conditions for $i(x,0)$ decay faster than $e^{-\lambda_0 x}$, λ_0 being the decay constant at $c = c_0$, then it can be shown that the selected speed for large t is c_0 (Aronson and Weinberger, 1978),

$$c_0 = 2\sqrt{d(R_0(R+S)-1)}. \quad (3)$$

This value of c also determines the decay constant at the front end of the wave, $\lambda_0 = c_0/(2d)$, which is the slowest decaying mode of any $c \geqslant c_0$.

We can estimate $d \approx 4\times10^{-7}$–5×10^{-7} by equating the value of 1 variant per year to $c_0 a(v+\mu)$, and setting $R_0 = 3$, $R + S \approx 0.99$ and $a \approx 7$. It follows that $D = d(v+\mu)a^2 \approx 1\times10^{-3} - 1.5\times10^{-3}$/year. These values of D appear to be comparable to the value of $D = p/2$ obtained from the steady mutation rate of 9.7×10^{-3} codon substitutions/year (substitutions causing amino acid replacements) in the HA1 subunit of the HA protein (Fitch et al., 1997). This is the region responsible

Table 1

Parameter	Symbol	Estimate	Source
Reproduction number	R_0	2–5	Spicer and Lawrence (1984)
Infectious period	$1/v$	2–10 days	Douglas (1975), Frank et al. (1981)
Cross-immunity scale	a	6–9	Potter et al. (1977), Pease (1987)

for most of the amino acid substitutions directly involved in antigenic drift. In doing the comparison, we must rely on the evidence provided by gene sequencing and cluster analysis (Wilson and Cox, 1990; Plotkin et al., 2002), suggesting that new drift strains require approximately four or more amino acid changes across two or more epitopes to successfully trigger an epidemic. Therefore, the estimate from sequence analysis is $D \approx 0.8 \times 10^{-3}$–$1.2 \times 10^{-3}$/year, which is of the same order of magnitude as the value obtained from the wave speed.

3. Threshold conditions and wave structure

To demonstrate the existence of a threshold condition (Anderson and May, 1991), we first observe that $(S, i, r) = (1, 0, 0)$ is an equilibrium state of system (2). The stability of the endemic equilibrium can be determined by studying perturbations around equilibrium of the form $(\sigma(t), j(x,t), \rho(x,t))$. The perturbations obey the equations

$$\frac{d\sigma}{dt} = -e\sigma(t) - R_0 J(t), \tag{4a}$$

$$\frac{\partial j(x,t)}{\partial t} = R_0 j(x,t) - j(x,t) + d\frac{\partial^2 j(x,t)}{\partial x^2}, \tag{4b}$$

where $J(t) = \int_{-\infty}^{\infty} j(x,t)\,dx$. Integrating the j-equation over x yields

$$\frac{dJ(t)}{dt} = R_0 J(t) - J(t) \tag{5}$$

showing that $J(t)$ grows iff $R_0 > 1$.

In the following, we provide some necessary conditions for the existence of traveling waves. Assume that there exists a steady traveling wave of the form $(S^*, i(u), r(u))$, where the coordinate $u = x - ct$ moves with the front and S^* is time independent. The wave solutions will have to solve the equations

$$0 = e(1 - S^*) - R_0 S^* I, \tag{6a}$$

$$-c\frac{di(u)}{du} = R_0 i(u)\int_{-\infty}^{u} K(u-w)r(w,t)\,dw + R_0 S^* i(u) - i(u) + d\frac{d^2 i(u)}{du^2}, \tag{6b}$$

$$-c\frac{dr(u)}{du} = -R_0 r(u)\int_{u}^{\infty} K(w-u)i(w)\,dw + (1-e)i(u) - er(u). \tag{6c}$$

From these equations, we must find a homoclinic orbit in the 3D phase space $(i(u), di(u)/du, r(u))$ that connects the $(0,0,0)$ state at $u = -\infty$ to the same state at $u = \infty$ (Kuznetsov, 1998). As $u \to -\infty$, the equations for i and r reduce to

$$-c\frac{di(u)}{du} = R_0 S^* i(u) - i(u) + d\frac{d^2 i(u)}{du^2}, \tag{7a}$$

$$-c\frac{dr(u)}{du} = -R_0 r(u) I + (1-e)i(u) - er(u) \tag{7b}$$

or into a set of three first-order equations

$$\frac{di(u)}{du} = y(u), \tag{8a}$$

$$\frac{dy(u)}{du} = [-cy(u) - R_0 i(u) S^* + i(u)]/d, \tag{8b}$$

$$\frac{dr(u)}{du} = [R_0 r(u) I - (1-e)i(u) + er(u)]/c. \tag{8c}$$

These linear equations have two positive real eigenvalues about $(0,0,0)$ provided that $1 - R_0 S^* > 0$. Assuming that $r > 0$ and integrating Eq. (6b) between $-\infty$ and $+\infty$, we find

$$0 = R_0 S^* I - I + R_0 \int_{-\infty}^{\infty} i(u)\int_{-\infty}^{u} K(u-w)r(w)\,dw\,du < R_0 I(S^* + R) - I, \tag{9}$$

which gives the inequalities $R_0 S^* - 1 < 0$ and $R_0(S^* + R) - 1 > 0$. This shows that we do in fact have two positive real eigenvalues.

As $u \to +\infty$, system (6) simplifies to

$$\frac{di(u)}{du} = y(u), \tag{10a}$$

$$\frac{dy(u)}{du} = [-cy(u) - R_0 i(u)(S^* + R) + i(u)]/d, \tag{10b}$$

$$\frac{dr(u)}{du} = [-(1-e)i(u) + er(u)]/c, \tag{10c}$$

which has two negative real eigenvalues provided that

$$c^2/d + 4(1 - R_0(S^* + R))/d > 0, \tag{11a}$$

$$1 - R_0(S^* + R) > 0, \tag{11b}$$

$$c > 0. \tag{11c}$$

Condition (11b) is satisfied according to the computation given above. Condition (11a)+(11c) holds for all $c \geqslant c_0$, where c_0 is the asymptotic speed we have given in the previous section. This analysis explains why we end up with the traveling pulse wave structure. Since at $u = -\infty$, we start with a 2D unstable manifold in a 3D space and since we end up at $u = +\infty$ with a 2D stable manifold these manifolds will intersect in the generic sense, giving us the homoclinic orbit we are looking for.

In Fig. 1, we compare the asymptotic speed (3) with numerical solutions for different values of R_0. The fit is close, within a 4–8% difference.

The reason why the numerical solutions always lie below the theoretical prediction has to do with system (1) being structurally unstable (instability to small

changes in the model and not to perturbations of the solutions). This structural instability arises from the instability of the homoclinic orbit to generic perturbations of the model (Kuznetsov, 1998). Two of the consequences of this instability are the slow convergence to the asymptotic wave and the strong dependence of the speed of the wave on either small random fluctuations in the parameters of the system or on the presence of a small cutoff ahead of the wave front (Paquette et al., 1994; Brunet and Derrida, 1997; Kessler et al., 1998). If one thinks of the continuous system (1) as an approximation to a contact process among N individuals, the cutoff is $1/N$. Both, the $1/t$ convergence to the asymptotic speed and the effective cutoff to the infection process introduced by localized initial conditions (we use rectangular distributions for $i(x,0)$ and $r(x,0)$) during numerical simulations for i and r reduce the velocity of the front.

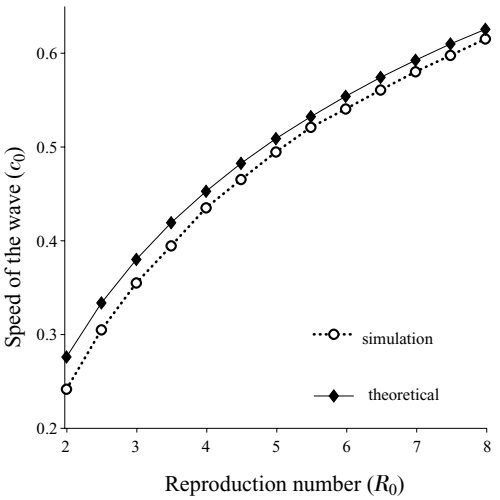

Fig. 1. Comparison between the theoretical speed of the infective wave (Eq. (3)) and the speed estimated by simulating the basic model (1) as functions of the reproduction number R_0. To simulate the basic model, we discretized each of the PDEs of system (1) with a set of ñ ODEs by equally spacing the variant interval (x_{min}, x_{max}) (numerical values are ñ $= 800$, $x_{min} = -10$, and $x_{max} = 150$). The speed in the simulation is evaluated as the slope of the straight line that best fits (in a least square sense) the average of the normalized distribution for the i's (i.e. $\int_{-\infty}^{\infty} x \cdot i_n(x)\,dx$) as a function of time. The algorithm used to integrate the ODEs system is a fully variable step size method (ODE113 described in Shampine and Reichelt, 1997). The convolution integral of Eq. (2b) has been evaluated via the Richardson's extrapolation of the Simpson rule (see formula 4.2.4 in Press et al., 1988). Parameter values are set to $\mu = 2.5 \times 10^{-4}$, $N = 1$, $v = 1$, $b = R_0(\mu + v) \cong R_0$, $K(z) = z/(z + a)$, $a = 5$ and $D = 0.02$.

Apart from these numerical issues, the presence of a real cutoff equal to $1/N$ in actual populations of N individuals can be estimated. Our model falls within the category of equations from which there are theoretical estimates (Brunet and Derrida, 1997, 1999). In the presence of a cutoff, the deterministic model has a velocity

$$c_N \simeq c_0 - \frac{(\pi\lambda_0)^2 c''(\lambda_0)}{2\ln(N)^2}. \tag{12}$$

In the last equation, $\lambda_0 = c_0/(2d)$ and the second derivative c'' is evaluated at this point. For example, if we assume $N \sim 10^6$ and the parameter values of Fig. 1, we find c_N to be $\approx 2.5\%$ smaller than c_0.

4. Asymptotic solutions

We are not aware of any exact solutions of system (2) and consequently, will seek asymptotic solutions based on the smallness of d. In Fig. 2, we display several snapshots of wave profiles as a function of R_0.

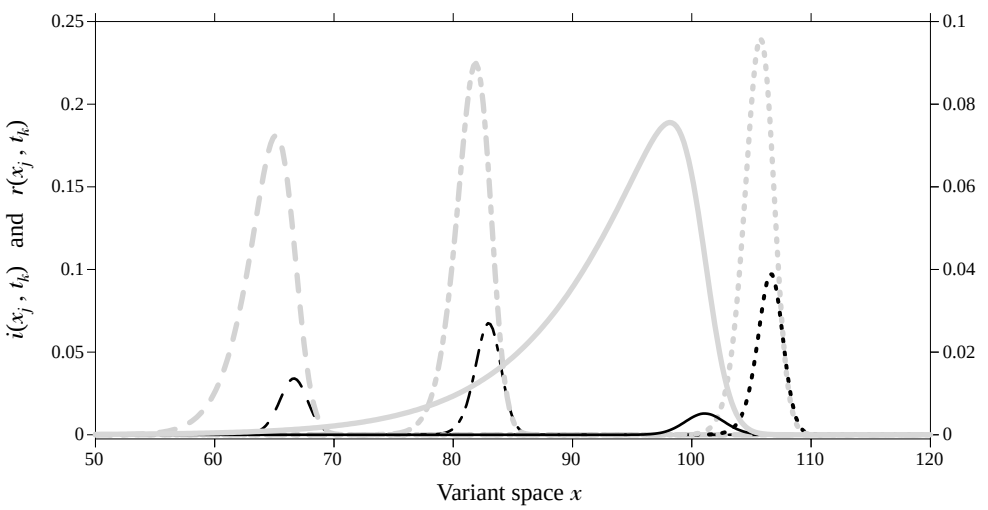

Fig. 2. Traveling waves for i (black) and r (light gray) obtained by simulating the system (1) for different values of the contact rate $b \cong R_0$: $R_0 = 2$ (solid, scale on the right y-axis), $R_0 = 4$ (dashed), $R_0 = 6$ (dashed and dotted), and $R_0 = 8$ (dotted). The snaphosts have been taken at times $T = 500$ for $R_0 = 2$, and $T = 240$ for all other R_0. Unspecified parameters and numerical scheme as in Fig. 1.

We observe that at low values of R_0, the distribution of recovered is highly skewed. When an epidemic outbreak is weak, the infected are unable to pull all recovered near its own distribution, thus leaving behind a significant trail of uninfected across all x. This skewness decreases significantly as R_0 increases. In contrast, the distribution of infected is highly symmetrical for all R_0. This is a consequence of the smoothing effect of diffusion along both directions. In Fig. 3, we plot the variance of the normalized distributions, $i_n(x) = i(x)/I$ and $r_n(x) = r(x)/R$, as functions of d. We observe that the variance of the i distribution scales reasonably well with $\sqrt{d}$, whereas the variance of the r distribution scales closer to $d^{1/4}$. We will explain this scaling later.

The results given above justify using an expansion for i of the exponential type $i(u) = e^{F(u)/\sqrt{d}}$ (geometrical optics approximation, Bender and Orszag, 1999). This is the sort of global approximation we seek out while solving second-order differential equations with a small parameter multiplying the highest derivative. The more precise physical optics or WKB approximation could also be used, but the results will not be altered in a significant way due to the constraints arising from the conservation of the total population and the slow variation in u of the term $F_0(u)$ in the next order expansion, $i(u) = e^{F(u)/\sqrt{d}+F_0(u)}$. In the frame of reference moving with the wave, $u = x - c_0 t$, system (2) is equivalent to (see also Eq. (6))

$$S^* = \frac{e}{e + R_0 I}, \tag{13a}$$

$$-c_0 \frac{di(u)}{du} = i(u) R_0 \left(\int_{-\infty}^{u} K(u-w) r(w)\, dw + S^* \right) - i(u) + d \frac{d^2 i(u)}{du^2}, \tag{13b}$$

$$-c_0 \frac{dr(u)}{du} = -r(u) R_0 \int_{u}^{\infty} K(w-u) i(w)\, dw + (1-e) i(u) - e r(u). \tag{13c}$$

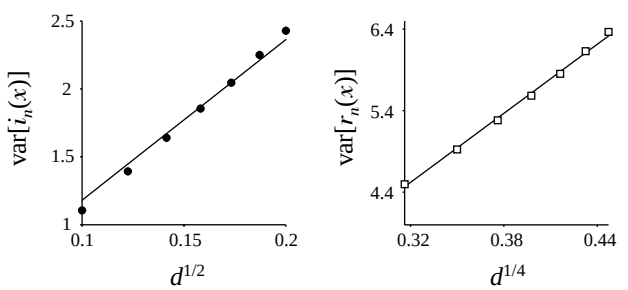

Fig. 3. The variance of i (filled circles) and the variance of r (white squares) as functions of $d^{1/2}$ and $d^{1/4}$, respectively. The best fit lines (solid) have the constraint of passing through the origin. $R_0 = 4$, and other parameter values as in Fig. 1.

The first derivative term of Eq. (13b) can be eliminated with the transformation

$$i(u) = e^{-\lambda_0 u} i_1(u), \tag{14}$$

where $\lambda_0 = c_0/(2d)$ and $c_0 = 2\sqrt{d(R_0(R+S^*)-1)}$, giving

$$d \frac{d^2 i_1(u)}{du^2} - i_1(u) R_0 \left(R - \int_{-\infty}^{u} K(u-w) r(w)\, dw \right) = 0. \tag{15}$$

The decay constant λ_0 scales as $d^{-1/2}$. Therefore, we write $i_1(u) = e^{F_1(u)/\sqrt{d}}$ and substitute this expression into Eq. (15). To lowest order in the expansion parameter d we obtain

$$\left(\frac{dF_1(u)}{du} \right)^2 - R_0 \left(R - \int_{-\infty}^{u} K(u-w) r(w)\, dw \right) = 0,$$

which has a solution $F_1(u) = \int_{-\infty}^{u} \sqrt{Q(v)}\, dv$, with $Q(v) = \int_{-\infty}^{v} K(v-w) r(w)\, dw$ (Bender and Orszag, 1999). There is a yet undetermined constant in front of i_1. Its value will be fixed later through the conservation equation, $S^* + I + R = 1$.

When $d \ll 1$, the i and r distributions are narrow as indicated in Fig. 3. Furthermore, K changes slowly over the diffusion length scale, thus allowing us to approximate $Q(v) = RK(v)$ for $v \geqslant 0$. In doing so, we assume that the peak value of $r(w)$ occurs at $w = 0$. This can always be done as the system is translationally invariant. The same can be said about integrals involving i in the integrand. The final expression for i is then

$$i(u) = \begin{cases} \exp\left(-\dfrac{c_0 u}{2d} + \sqrt{\dfrac{R_0 R}{d}}\, u \right) & \text{if } u < 0, \\ \exp\left(-\dfrac{c_0 u}{2d} + \displaystyle\int_{-\infty}^{u} \sqrt{\dfrac{R_0 R(1-K(v))}{d}} \right) & \\ \quad \text{otherwise.} & \end{cases} \tag{16}$$

Note that the solution above has the correct asymptotic behavior as $u \to \pm\infty$. In the discussion that follows we will use the Monod kernel, $K(v) = v/(1+v)$. We expect other smooth kernels to exhibit similar qualitative behavior.

In the following discussion, we take advantage of the symmetrical nature of the i distribution. We first find the root $\bar{u}$ of the equation $F(\bar{u}) = 0$ in the exponent of $i(u)$. This parameter measures the separation between i and r maxima. For the Monod kernel we find

$$\bar{u} = \frac{1 - R_0 S^*}{R_0(1-I) - 1}. \tag{17}$$

In general, $S^* = e/(e + R_0 I) \approx 1/(R_0 n_0) \ll 1$, where n_0 is the average number of reinfections in a host lifetime and $I \ll R$, which means that $\bar{u} \approx 1/(R_0 - 1)$ for $R_0 \geqslant 2$. Fig. 4 shows that the distance between peaks in unscaled units

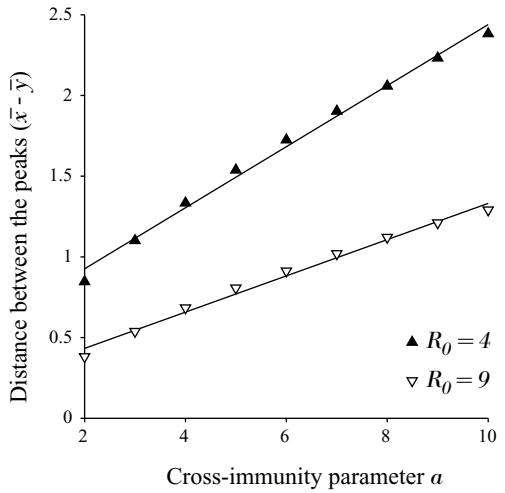

Fig. 4. Distance between the peaks of i and r as a function of the cross-immunity parameter a. The location of the peaks of the i distribution ($\bar{x}$) and of the r distribution ($\bar{y}$) on the variant space x is evaluated as the maximum of the third-order polynomials that best fits the six closest neighbors to the peak of the discretized distributions. Parameter values as in Fig. 1.

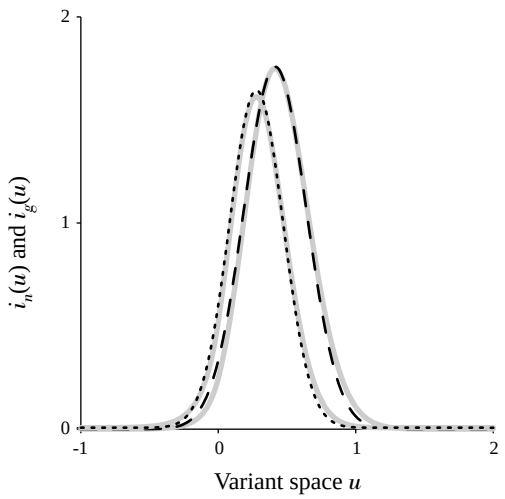

Fig. 5. The gaussian approximation for i (black) compared to asymptotic i (gray) for two different values of R_0: $R_0 = 6$ (dotted) and $R_0 = 4$ (dashed). Parameter values as in Fig. 1.

is indeed a linear function of a. The prediction agrees well with these values.

The normalized gaussian approximation to i can then be written as

$$i_g(u) = \frac{\exp(-(u-\bar{u})^2/2\,\mathrm{var}(\bar{u}))}{\sqrt{2\pi\,\mathrm{var}(\bar{u})}}$$

with $var(\bar{u}) = \sqrt{d}/|F''(\bar{u})|$. In Fig. 5, we compare the gaussian approximation to $i_n(u)$. The fit is quite good for all values of R_0 (slightly worse at larger values) and it gets better at smaller values of d. The gaussian form can be used to compute integrals involving $i(u)$ in the integrand. In addition, it shows that I scales approximately as $d^{1/4}$.

We now address the asymptotic solution for $r(u)$. The narrow i distribution simplifies the r equation to

$$c_0\frac{\mathrm{d}r(u)}{\mathrm{d}u} + r(u)(e + R_0 IK(\bar{u}-u)) + (1-e)Ii_n(u) = 0, \quad (18)$$

where $K(\bar{u}-u)$ vanishes if $u > \bar{u}$. For $u < \bar{u}$ we find

$$r_<(u) = \frac{I(1-e)}{2c_0}\exp\left(\frac{R_0 I}{c_0}\int_{\bar{u}}^{u} K(\bar{u}-v)\,\mathrm{d}v\right).$$

To derive this expression, we use the fact that the exponent of i scales as $d^{-1/2}$ but the exponent of the other integrand scales as $d^{-1/4}$, justifying taking the area of the fast component and evaluating the other factor at $\bar{u}$. The rather long left tail of $r(u)$ is caused by the $d^{-1/4}$ scaling and it explains why the variance of $r(u)$ scales closer to $d^{1/4}$ than to $\sqrt{d}$. The very small correction term e/c_0 also scales as $d^{-1/2}$ but since $e \sim 10^{-4}$, it has little effect on the final result. The factor $\frac{1}{2}$ multiplying I has to do with the lower limit of the integral ending at $\bar{u}$, thus including only half of the total number of infected (almost half within at most a few %, due to the symmetrical nature of $i(u)$). The solution r for $u > \bar{u}$ can be written as

$$r_>(u) = B\frac{I(1-e)}{c_0}\int_u^{\infty} \mathrm{e}^{e(u-v)/c_0} i_n(v)\,\mathrm{d}v.$$

In the asymptotic analysis of i, we locked the peak of the r distribution at $u = 0$, yet the solution we have just found for $r_<(u)$ has its peak value at $\bar{u}$. We translate this solution by replacing $u \to u + \bar{u}$ so that $r_<(u + \bar{u})$ has its maximum at $u = 0$. We then find B by matching the two r solutions at $u = 0$. The final $r(u)$ is in general piecewise continuous.

In Fig. 6, we display the exponential approximation and numerical curves. The fit is qualitatively correct with larger i's for the asymptotic estimates. We attribute this effect to the approximation $Q(v) = RK(v)$, which overestimates the integral in Eq. (15). These estimates improve as d gets smaller.

The scaling $I \sim d^{1/4}$ predicts a value of $I \approx 0.4$–2% ($R_0 \approx 2$–5, parameter values as in Fig. 3) when

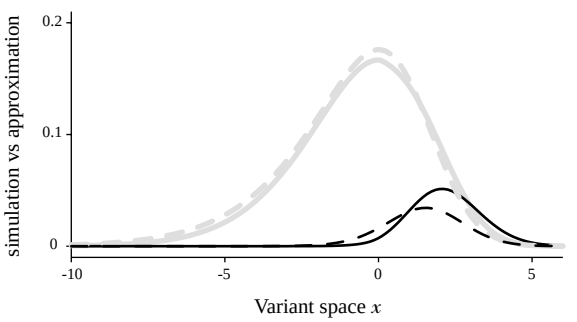

Fig. 6. Comparison of numerical (dashed) and asymptotic (solid) i (black) and r (gray). Except for $D = 0.02$ and $R_0 = 4$, other parameter values as in Fig. 1.

$d \approx 4 \times 10^{-7}$, which is above but of the same order of magnitude as those values reported in the literature (Fox et al., 1982; Monto and Kioumehr, 1975), $I_t \approx 0.15$–0.48%. At these levels of I, the total susceptible class S increases to a few percent. One final comment. To find the appropriate asymptotic amplitudes for i and r we must start with an initial trial value for I. We use this value of I to find the initial equilibrium value of S^* in Eq. (6a), and then obtain $R = 1 - I - S^*$. After we finish a first round of computations we end up with a distribution of recovered equal to $r(u)$, which we integrate to find R_t. We repeat this process several times with new trial values of I until $|R_t - R| \ll I$.

5. Discussion

The simultaneous cocirculation of many variants of the same subtype of influenza A in the human population poses enormous challenges to epidemiologists and theoretical biologists alike. If one keeps track of the extensive life history of infections in each host then the population has to be partitioned into a large number of dynamical classes giving rise to complex dynamics (Andreasen et al., 1997; Gupta et al., 1998; Lin et al., 1999). Yet, these models assume that the number of variants is fixed, giving no mechanism for the generation of new strains through drift. The simple continuous model we have analysed has the advantage that mutation enters naturally in the process of generating diversity while predicting a sequential evolution of the virus in antigenic space. Despite the shortcomings of the model—keeping track of only the last infection and crude cross-immunity structure—it is encouraging to find that the calculated speed of drift is of the right order of magnitude. This simple result may help us to understand drift as a function of different parameters, susceptible pool, mutation rate, susceptibility and reproduction number. Moreover, this result suggests that other fruitful insights on this complex disease may be forthcoming from the analysis of simple (discrete, continuous, or stochastic) models.

We have seen in Section 2, that the predicted speed of the wave compares well to data from numerical simulations. Yet, other considerations deserve attention. The drift process is probably more complex than the simple diffusion approximation we propose as amino acid substitutions usually occur in one or simultaneously in several different epitopes of HA1 to trigger an epidemic season (Both et al., 1983; Plotkin et al., 2002). There are substantial correlations at the antigenic level but sometimes even those changes do not guarantee a successful virus survival in the population. This survival may also depend on exogenous (seasonality, climate) or endogenous factors (immunity of the host).

Clearly much work remains to be done in characterizing what antigenic properties are important for virus reproduction within the host and what virus concentration threshold levels are needed before hosts become infectious. Neither have we addressed the spatial distribution of individuals in the population that is required before the virus can take off as a local or global epidemic outbreak. The simple threshold condition $R_0 > 1$, based on total population size N, is oblivious of the clustering and constant migration patterns observed in natural populations. A stochastic formulation is needed to resolve issues connected on the one hand, to the non-uniformity of amino acid cluster evolution at small scales (Plotkin et al., 2002), and on the other hand, to the extinction time of the disease in small to medium populations at large scales.

In Section 4, we demonstrated the usefulness of asymptotic estimates. Several parameters were found to possess simple scaling properties. We have found that the total fraction of infected I and the variance of $r(x)$ scales with mutation rate p approximately as $p^{1/4}$, that the variance of $i(x)$ and the speed of the wave scale as $p^{1/2}$ and, finally, that the distance between the i and r peaks scale approximately as $a/(R_0 - 1)$ for $R_0 \geqslant 2$, where a is the length scale of cross-immunity. The model predicts a constant population $I(t)$ over evolutionary time-scale, but it does not explain the temporal variation of $I(t)$ during annual epidemics. Perhaps this shortcoming can be addressed by developing a seasonally forced version of model (1). We hope to explore these and other issues in the near future.

The simple model we have just discussed is only a first step in understanding influenza A virus drift. Nevertheless it suggests avenues for extending, along several directions, our comprehension of how the complex patterns of virus–host coevolution shape the dynamic structure of herd immunity.

Acknowledgements

We are grateful to Jonathan Dushoff, Joshua Plotkin, Freddy Christiansen, David Earn and Lorenzo Codecasa for critical comments and useful discussions. This work has been supported in part by NIH Grant #1RO1GM607929.

References

Anderson, R.M., May, R.M., 1991. Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, Oxford, pp. 17–21.

Andreasen, V., Levin, S.A., Lin, J., 1996. A model of influenza A drift evolution. Z. Angew. Math. Mech. 76 (2), 421–424.

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

Aronson, D.J., Weinberger, H.F., 1978. Multidimensional nonlinear diffusion arising in population genetics. Adv. Math. 30, 33–76.

Bender, C.M., Orszag, S.A., 1999. Advanced Mathematical Methods for Scientists and Engineers. Springer, New York, pp. 484–494.

Billingham, J., King, A.C., 2000. Wave Motion. Cambridge University Press, Cambridge, pp. 343–345.

Both, G.W., Sleigh, M.J., Cox, N.J., Kendal, A.P., 1983. Antigenic drift in influenza virus H3 hemagglutinin from 1968 to 1980: multiple evolutionary pathways and sequential amino acid changes at key antigenic sites. J. Virol. 48, 52–60.

Brunet, E., Derrida, B., 1997. Shift in the velocity of a front due to a cutoff. Phys. Rev. E 56, 2597–2604.

Brunet, E., Derrida, B., 1999. Microscopic models of traveling wave equations. Comput. Phys. Commun. 121–122, 376–381.

Buonagurio, D.A., Nakada, S., Parvin, J.D., Krystal, M., Palese, P., Fitch, W.M., 1986. Evolution of human influenza A viruses over 50 years: rapid, uniform rate of change in NS gene. Science 232, 980–982.

Bush, R.M., Fitch, W.M., Bender, C.A., Cox, N.J., 1999. Positive selection on the H3 hemagglutinin gene of human influenza virus A. Mol. Biol. Evol. 16, 1457–1465.

Couch, R.B., Kasel, J.A., 1983. Immunity to influenza in man. Annu. Rev. Microbiol. 37, 529–549.

Cox, N.J., Subbarao, K., 2000. Global epidemiology of influenza: past and present. Annu. Rev. Med. 51, 407–421.

Douglas Jr., R.G., 1975. Influenza in man. In: Kilbourne, E.D. (Ed.), The Influenza Viruses and Influenza. Academic Press, New York, pp. 395–447.

Earn, D.J.D., Dushoff, J., Levin, S.A., 2002. Ecology and evolution of the flu. Trends Ecol. Evol. 17, 334–340.

Fitch, W.M., Leiter, J.M.E., Li, X., Palese, P., 1991. Positive Darwinian evolution in human influenza A viruses. Proc. Natl Acad. Sci. USA 88, 4270–4274.

Fitch, W.M., Bush, R.M., Bender, C., Cox, N.J., 1997. Long term trends in the evolution of H(3) HA1 human influenza type A. Proc. Natl Acad. Sci. USA 94, 7712–7718.

Fox, J.P., Hall, C.E., Cooney, M.K., Foy, H.M., 1982. Influenza virus infections in Seattle families, 1975–1979. I. Study design, methods and the occurrence of infections by time and age. Am. J. Epidemiol. 116, 212–227.

Frank, A.L., Taber, L.H., Wells, C.R., Wells, J.M., Glezen, W.P., Paredes, A., 1981. Patterns of shedding of myxoviruses and paramyxoviruses in children. J. Infect. Dis. 144 (5), 433–441.

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.

Gupta, S., Ferguson, N., Anderson, R., 1998. Chaos, persistence and evolution of strain structure in antigenically diverse infectious agents. Science 280, 912–915.

Kessler, D.A., Ner, Z., Sander, L.M., 1998. Front propagation: precursors, cutoffs, and structural stability. Phys. Rev. E 58, 107–114.

Kuznetsov, Y.A., 1998. Elements of Applied Bifurcation Theory. Springer, New York, pp. 195–248.

Larson, H.E., Tyrrell, D.A.J., Bowker, C.H., Potter, C.W., Schild, G.C., 1978. Immunity to challenge in volunteers vaccinated with an inactivated current or earlier strain of influenza A (H3N2). J. Hyg. Cambridge 80, 243–248.

Levin, S.A., Pimentel, D., 1981. Selection of intermediate rates of increase in parasite–hosts systems. Am. Nat. 117, 308–315.

Levine, A.T., 1992. Viruses. W. H. Freeman, New York, pp. 155–175.

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

Monto, A.S., Kioumehr, F., 1975. The Tecumseh study of respiratory illness. IX. Occurrence of influenza in the community, 1966–1971. Am. J. Epidemiol. 102, 553–563.

Murray, J.D., 1989. Mathematical Biology. Springer, New York, pp. 274–281.

Paquette, G.C., Chen, L.Y., Goldenfeld, N., Oono, Y., 1994. Structural stability and renormalization group for propagating fronts. Phys. Rev. Lett. 72, 76–79.

Pease, C.M., 1987. An evolutionary epidemiological mechanism, with applications to type A influenza. Theor. Popul. Biol. 31, 422–452.

Plotkin, J.B., Dushoff, J., Levin, S.A., 2002. Hemagglutinin sequence clusters and the antigenic evolution of influenza A virus. Proc. Natl Acad. Sci. USA 99, 6263–6268.

Potter, C.W., Jennings, R., Nicholson, K.A.J., Tyrrell, D., Dickinson, K.G., 1977. Immunity to attenuated influenza virus WRL 105 infection induced by heterologous, inactivated influenza A virus vaccines. J. Hyg. Cambridge 79, 321–332.

Press, W.H., Flannery, B.P., Teukolsky, S.A., Vettering, W.T., 1988. Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press, Cambridge, UK.

Sasaki, A., 1994. Evolution of antigenic drift and switching: continuously evading pathogens. J. theor. Biol. 168, 291–308.

Shampine, L.F., Reichelt, M.W., 1997. The Matlab ODE Suite. SIAM J. Sci. Comput. 18, 1–22.

Smith, F.L., Palese, P., 1989. Variation in influenza virus genes. In: Krug, R.M., Fraenkel-Conrat, H., Wagner, R.R. (Eds.), The Influenza Viruses. Plenum, New York, pp. 319–359.

Spicer, C.C., Lawrence, C.J., 1984. Epidemic influenza in Greater London. J. Hyg. Cambridge 93, 105–112.

Webster, R.G., Bean, W.J., Gorman, O.T., Chambers, T.M., Kawaoka, Y., 1992. Evolution and ecology of influenza A viruses. Microbiol. Rev. 56, 152–179.

Wilson, I.A., Cox, N.J., 1990. Structural basis of immune recognition of influenza virus hemagglutinin. Annu. Rev. Immunol. 8, 737–771.

Xu, X., Rocha, E.P., Regenery, H.L., Kendal, A.P., Cox, N.J., 1993. Genetic and antigenic analyses of influenza A (H1N1) viruses, 1986–1991. Virus Res. 28, 37–55.