

Shaping the phylogenetic tree of influenza by cross-immunity

Viggo Andreasen^{a,*}, Akira Sasaki^b

^a*Department of Mathematics and Physics, Roskilde University, DK-4000 Roskilde, Denmark*

^b*Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812-81, Japan*

Received 15 February 2005

Available online 27 April 2006

Abstract

Cross-immunity among related strains can account for the selection producing the slender phylogenetic tree of influenza A and B in humans. Using a model of seasonal influenza epidemics with drift (Andreasen, 2003. Dynamics of annual influenza A epidemics with immuno-selection. *J. Math. Biol.* 46, 504–536), and assuming that two mutants arrive in the host population sequentially, we determine the threshold condition for the establishment of the second mutant in the presence of partial cross-protection caused by the first mutant and their common ancestors. For fixed levels of cross-protection, the chance that the second mutant establishes increases with ρ the basic reproduction ratio and some temporary immunity may be necessary to explain the slenderness of flu's phylogenetic tree. In the presence of moderate levels of temporary immunity, an asymmetric situation can arise in the season after the two mutants were introduced and established: if the offspring of the new mutant arrives before the offspring of the resident type, then the mutant-line may produce a massive epidemic suppressing the original lineage. However, if the original lineage arrives first then both strains may establish and the phylogenetic tree may bifurcate.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Drift evolution; Herd immunity; Epidemiology; Antigenic drift; Natural selection; Haploid

1. Introduction

The phylogenetic trees of influenza genes are long and narrow with short side branches and rare bifurcations where two trunks coexist over several years (Buonagurio et al., 1986; Cox and Subbarao, 2000; Fitch et al., 1991, 1997; Hay et al., 2001). We here show how cross-immunity among genetically related viral variants can allow one mutant to establish while suppressing other equally viable mutants of the same ancestor, a mechanism that will give rise to a narrow tree with short side branches.

The slender phylogenetic tree of influenza A contrasts with the phylogenetic trees that have been observed for influenza in birds and for genes of other infectious diseases (Kawaoka et al., 1998; Frey et al., 1998; Rambaut et al., 2001). From a theoretical view-point a virus species can be seen as a quasi-species, i.e. as a cloud of viral types kept

together in a balance between mutation and selection (Eigen and Schuster, 1979; Eigen, 1993; Abu-Raddad and Ferguson, 2004). For such a system we would expect a constant rate of divergence in non-structural genes suggesting a more branched structure of the phylogenetic tree than that observed for the flu. The mechanism thought to be responsible for the shape of the flu tree is natural selection induced by the cross-immunity among related viral strains. Such cross-immunity constantly inhibits branching by natural selection against mutants that are related to previously successful variants allowing only one lineage to spread in the population (Buonagurio et al., 1986).

Influenza A is an example of a virus undergoing antigenic change at an intermediate time scale, longer than the duration of an infection event, yet shorter than host life span. Thus during an infection, a single antigenic strain of the pathogen colonizes a host, but through its life the same host may be infected several times by antigenically different variants. The selection processes in the viral population giving rise to antigenic change and to

*Corresponding author. Fax: +45 4674 3020.

E-mail addresses: viggo@ruc.dk (V. Andreasen),
asasascb@mbox.nc.kyushu-u.ac.jp (A. Sasaki).

the exclusion of multiple lineages are therefore determined by the immunity structure of the host population (Grenfell et al., 2004).

The most important antigen of influenza A and B is the hemagglutinin surface-molecule (HA), although antibodies are formed in response to many other sites, most noticeably the neuraminidase antigen (NA). Due to its configuration, antibodies are not formed to the functionally active part of the HA-molecule, instead antibodies are formed to five non-functional epitopes allowing for significant diversity in flu antigens. The antigenic variation in influenza virus is caused by two distinct processes. In a process known as virus drift, point mutations in the gene coding for HA give rise to new virus variants (strains) with gradually changing antigenic properties. Immunity obtained from infection with a specific strain of influenza confers permanent immunity to that particular strain and a partial protection against related strains. In general, the level of cross-protection decreases with the number of amino acids by which the HA-gene of the two strains differ and hence with the distance between the two strains in Hamming space. Smith et al. (2002) suggest that 3–4 amino-acid substitutions must occur in the virus before there is an appreciable chance of reinfection of the same host and in general cross-immunity seems to protect against the mutations that accumulate over a few years (Smith et al., 2002; Cox and Subbarao, 2000; Potter et al., 1977; Larson et al., 1978). Recently Smith et al. (2004) have found that the antigenic variation in influenza A/H3N2 can be represented in a two-dimensional space with most of the variation occurring along a single axis.

In addition to antigenic drift antigens also change in distinct *shifts* where reassortment with avian influenza replaces whole segments of the viral surface-structure introducing a new *subtype*. Such shift events occur at irregular intervals on the order of decades and they are usually associated with the disappearance of the old subtype. Thus, our focus will be on an intermediate time scale describing the period between two shifts.

Previous models of influenza drift have focused on the epidemiological consequences of drift by assuming that mutations occur along a one-dimensional axis representing the main trunk of the phylogenetic tree and that the drift-mutation is constant over time (Pease, 1987; Inaba, 2001; Thieme and Yang, 2002; Girvan et al., 2002; Andreasen, 2003). They have modeled the speed of the drift along the axis (Andreasen et al., 1996; Gog and Grenfell, 2002; Lin et al., 2003) or the mutation rate required for viral drift (Boni et al., 2004). Because of the multiple strains and the complexities of the population-based herd-immunity, an account for the transmission dynamics and mutation process sufficiently detailed to reproduce the drift-like behavior seems to require individual-based computer simulations (Ferguson et al., 2003; Tria et al., 2005). To avoid the complexities of such models we will not attempt to include all processes involved in influenza drift but rather study branching as a “perturbation” of the normal

drift process. Most of the analytical drift-models do not allow for the introduction of additional mutants because the immunological structure of the virus population is modeled in a way that links to the time-progression so we shall base our model on that of Andreasen (2003).

2. Derivation of basic model

The basic idea in Andreasen’s drift-model is to separate completely the time scale of the epidemic from that of the drift process. Thus at the beginning of a season, a drift variant is introduced into the population. If sufficient susceptibility is present in the population, an epidemic occurs and irrespective of whether an epidemic occurred or not, the strain disappears at the end of the season. At the onset of the subsequent season a new drift variant appears. Based on the outcomes of the previous epidemics, the susceptibility to this strain is determined and a new epidemic may arise.

Mutation will not be described explicitly. Rather we assume that at the end of each season sufficient genetic variation is present to ensure that at least one new strain will establish in the subsequent season. The mutant strains appear towards the end of the flu-season, circulate at low, possibly decreasing, numbers during the low-transmission period and start their exponential growth phase in the next high-transmission season. If all strains occur at equally low prevalence at the beginning of the season, the strain with the highest growth rate will be the first epidemic strain and we will assume that this epidemic will run to its conclusion prior to the appearance of the next strain, “*sequential epidemics*”. From a modeling view point the assumption of sequential epidemics is an approximation that is valid for small inoculum size $I(0) \ll 1$ and thus a natural consequence of the time scale separation. As indicated in Fig. 1, a difference in the growth-rates of the two strains implies that the fastest growing strain almost completes its epidemic prior to a second strain attaining detectable levels provided that the initial size of the infectious population is small (compare to Ohtsuki and Sasaki, 2006; Gog et al., 2003).

From a biological view point some caution is called for. Specific strains appear to be well suited for the low-transmission period (Gog et al., 2003) and in addition genetic and antigenic variation is always present at least at the global and regional scales (Plotkin et al., 2002; Holmes et al., 2005). Still within a single epidemic, variation appears to be limited (Schweiger et al., 2002; Shih et al., 2005). We shall return to these issues in the discussion but for now we ignore such complications and assume the strains arrive to the host population sequentially as described. This will establish a baseline to which one can compare the more complex situation where cocirculation occurs (Boni et al., 2006).

Other strains circulating at the onset of the season, and hence the second strain appearing in the host population, are likely to have a lower growth rate or possibly higher

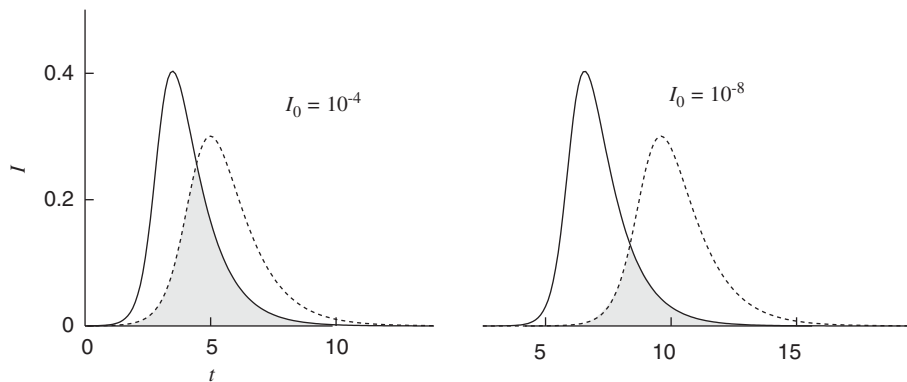


Fig. 1. The overlap of two independent (*SIR*) epidemics with different basic reproduction number starting from low numbers. The assumption of sequential epidemics corresponds to neglecting the interactions occurring in the shaded area. The approximation of sequential epidemics works well for small size of the initial infectious population. Parameter values used: $\rho_A = 4$ (full line) and $\rho_B = 3$ (broken line) while the infectious period is 1 and population size is set to 1. In both graphs initial conditions are $S = 1 - I_0$, $R = 0$, and $I_A = I_B = I_0$ as indicated on the figure.

cross-immunity to previous strains than do the first successful strain, or they would have been the first successful strain! To avoid parameter proliferation, we focus on an upper bound overestimating the growth potential of the second strain. Thus, we assume that the second strain is as fit as the first epidemic strain in the sense that the two strains have the same reproduction number in a fully susceptible population and both exhibit the same cross-reactivity towards their common ancestors. Since the two mutants are supposed to arise through two independent mutations we will assume that they differ by twice the amount of the annual drift thus eliciting a cross-reaction between them of the same magnitude as that produced towards the strain occurring two seasons ago. If both strains can cause an epidemic in the host population, a bifurcation in the phylogenetic tree has occurred and as the two lineages separate further in subsequent seasons, persistence of both branches is expected. Since we determine the branching conditions for a strain that is more fit than most mutants, our results should be interpreted probabilistically in the sense that the more easily the conditions for branching are met, the more likely it is that a strain with the necessary properties would arise. In our model derivation we will assume that single annual strain replacement happens in most years and hence we implicitly assume that bifurcations are rare events, so for parameter values well within the region where bifurcations can occur the model will no longer hold. For such parameter values it is an open question if the genetic structure of the virus is still characterized by drift at all (cf. Abu-Raddad and Ferguson, 2004, 2005).

Since our analysis focuses on the situation where branching is a rare event we will start from a situation where annual strain replacement by drift has settled at equilibrium and we now sketch a simplified version of the drift-model by Andreassen (2003) describing this situation. For details of the analysis see Andreassen (2003). We first study the dynamics of the epidemic within a single season.

Assuming that the level of cross-immunity depends only on the most related previous infection (which for now is

identical to the most recent infection), the immune structure of the host population at the beginning of a flu-season is described by

$$s_k, \quad k = 1, \dots \quad (1)$$

giving the fraction of the host population that was most recently infected k seasons ago. Since drift is supposed to occur at a constant speed, we identify the difference in years between the occurrence of two strains with distance in nucleotide composition and with the distance in antigenic properties so that s_k , $k = 1, \dots$ in fact summarizes the herd-immunity structure of the population. To keep the model simple we will assume that the population is closed in the sense that population size is constant and that no migration, births, and deaths occur. For mathematical convenience we will assume that infinitely many flu-seasons have occurred allowing the index in (1) to run to infinity. In addition, we will assume that cross-immunity acts by reducing *infectivity* rather than *susceptibility* because this leads to a significantly simpler model (Ferguson and Andreassen, 2002). These simplifying assumptions can be altered without qualitatively changing our conclusions and we return to the issue in the discussion.

At the onset of the epidemic season we assume that a few infected hosts come to the population from an external source. Let $S_k(t)$ denote the fraction of hosts whose most recent infection occurred k seasons ago and who in this season have not yet been infected at time t , while $I_k(t)$ denotes the fraction of hosts that are currently infected and whose last previous infection occurred k seasons ago. We will not need to keep track of those hosts that have recovered from infection. The course of the epidemic during the season in question now evolves according to a mass-action model

$$\begin{aligned} \dot{S}_k &= -\Lambda S_k, \\ \dot{I}_k &= \Lambda S_k - \nu I_k, \end{aligned} \quad (2)$$

where $\Lambda = c \sum \tau_k I_k$ is the force of infection and τ_k gives the reduction in infectivity due to acquired immunity. The parameter ν gives the rate of recovery from infection while c measures the contact rate in absence of any cross-immunity. Initial conditions are $S_k(0) = s_k, k = 1, \dots$ and $0 < \Lambda(0) \ll 1$. It turns out that we need not describe distribution of cross-immunities of the inoculum.

The model for the epidemic season can be simplified by using $Q = \sum \tau_k S_k$ and Λ as a dynamic variables. We will refer to Q as the potential infectivity because Q measures how much infectivity could be produced if all hosts become infected. With these new variables the course of the epidemic is determined by

$$\dot{Q} = \sum \tau_k \dot{S}_k = -\Lambda Q, \tag{3}$$

$$\dot{\Lambda} = \rho \sum \tau_k \dot{I}_k = \rho \sum \tau_k S_k \Lambda - \Lambda = \rho Q \Lambda - \Lambda, \tag{4}$$

where in addition time is rescaled in units of average duration of infection $1/\nu$ and $\rho = c/\nu$ is the basic reproduction number.

Clearly model (3) and (4) is the classical model of Kermack and McKendrick (1927). Since

$$\frac{d\Lambda}{dQ} = -\rho + \frac{1}{Q},$$

we can determine the outcome of the epidemic in terms of $\phi = Q(\infty)/Q(0)$ by observing that

$$0 = \Lambda(\infty) - \Lambda(0) = \log \phi + \rho q(1 - \phi), \tag{5}$$

with $q = Q(0)$. For the individual immunity classes S_k , we observe that $dS_k/dQ = S_k/Q$ so that the fraction of hosts in immunity class k that avoid infection during the entire epidemic is

$$S_k(\infty) = \phi S_k(0).$$

It is well known that Eq. (5) has exactly one solution $0 < \phi < 1$ if $\rho q > 1$ and none if $\rho q < 1$. In addition $\phi = 1$ is always a solution. Biologically $\phi = 1$ corresponds to the situation where no epidemic occurs and since we have assumed that an epidemic occurs when possible we exclude the solution at $\phi = 1$ when $\rho q > 1$. If no epidemic is possible, there will be no growth in the virus population and hence no new genetic variation allowing the drift process to continue and we will assume that the virus lineage will die out.

We have now determined how the immune structure of the host population changes from the onset of the epidemic to the end of the epidemic. According to our separation of the epidemic and drift processes a new strain appears prior to the next season. Assuming that the reproduction number ρ and cross-reactions τ_k remain unchanged from season to season, we can now determine how the immune structure changes from the onset of the T th season to the onset of the $(T + 1)$ st season.

At the onset of season $T + 1$ the immunity class s_1 consists of those hosts who were infected during the

previous season so that

$$s_1^{T+1} = \sum_k (1 - \phi) s_k^T = 1 - \phi$$

while those hosts that will enter immunity class $s_k, k = 2, \dots$ at the onset of the $(T + 1)$ th season will be those hosts who were in the s_{k-1} immunity class at the beginning of season T and who did not get infected, i.e.

$$s_k^{T+1} = \phi s_{k-1}^T, \quad k = 2, \dots$$

This completes the formulation of the model which can now be expressed in terms of a discrete map linking the immunity structure at the onset of season T to the immune-structure at the beginning of season $T + 1$:

$$(s_k^T)_{k=1\dots} \mapsto (s_k^{T+1})_{k=1\dots},$$

see Andreasen (2003) for details.

To simplify this map we will follow Boni et al. (2004) and make one additional assumption about the infectivity reduction τ_k allowing us to obtain a Markov-like property for the potential infectivity. Although cross-reactions are well documented in hemagglutination studies, the relation between the cross-reactions observed *in vitro* and the cross-immunity expressed in terms of the infectivity reduction factors τ_k remains unclear. Clearly cross-immunity should decline as strains become more dissimilar corresponding to τ_k being an increasing function of k reaching insignificant levels of cross-protection for strains that are a few seasons apart. We will assume that cross-immunity decays geometrically such that

$$\tau_k = 1 - \alpha^k,$$

where $\alpha < 1$ is a parameter describing how far cross-immunity reaches.

The specific infectivity-reduction factors we have chosen, satisfies the recursion formula

$$\tau_{k+1} = \tau_1 + \tau_k - \tau_1 \tau_k.$$

This relation allows us to capture the entire immunity structure at the onset of the epidemic season in a single value q , since we have that the potential infectivity at the start of the next season is

$$\begin{aligned} q^{T+1} &= \sum_1^\infty \tau_k s_k^{T+1} \\ &= \tau_1(1 - \phi) + \sum_1^\infty \tau_{k+1} \phi s_k^T \\ &= \tau_1(1 - \phi) + \phi \sum_1^\infty (\tau_1 + \tau_k - \tau_1 \tau_k) s_k^T \\ &= \tau_1(1 - \phi) + \tau_1 \phi \sum_1^\infty s_k^T + \phi(1 - \tau_1) \sum_1^\infty \tau_k s_k^T \\ &= \tau_1 + \phi(1 - \tau_1) q^T \\ &= 1 - \alpha + \alpha \phi q^T, \end{aligned}$$

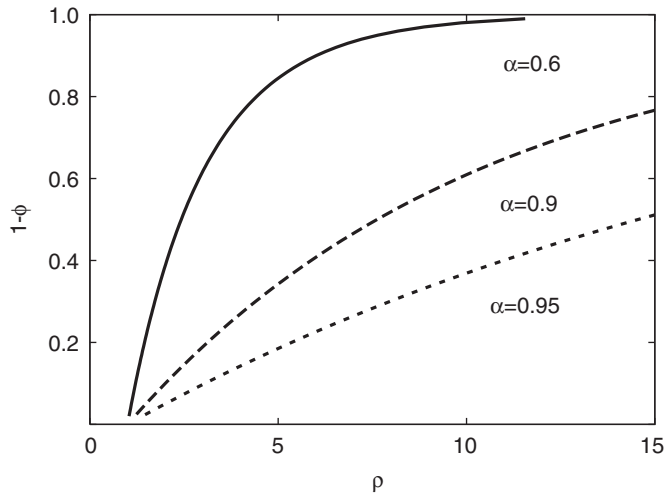


Fig. 2. The “attack rate” $1 - \phi$ i.e. the fraction of the host population that gets infected during an epidemic season when the drift process is at equilibrium. Here ρ denotes the basic reproduction number while α describes how far cross-immunity reaches in that $1 - \alpha^k$ gives the infectivity if infected k seasons after last infection relative to the infectivity of an immunologically naive host.

where $\phi = \phi^T$ denotes the fraction of hosts that escape infection during season T .

Expressing our model of the season-to-season dynamics in terms of the potential infectivity at the onset of the epidemic season now yields a one-dimensional model

$$q_{t+1} = 1 - \alpha + \phi \alpha q_t, \tag{6}$$

where $\phi = \phi(q)$ is the solution to (5). In the appendix we show that for $\rho > 1$ this model has a unique stable equilibrium corresponding to a situation where an epidemic of the same size will arise in every season. Fig. 2 shows how the “attack rate,” i.e. the fraction of the host population that is infected during a single epidemic, at equilibrium $1 - \phi$ depends on ρ and α . As expected the attack rate increases with the reproduction number ρ and decreases with the duration of cross-protection α .

3. Selection in a drifting virus population

We now turn our attention to the conditions that will meet a second mutant strain when entering the population in a given season. We will assume that drift has already occurred for sufficiently long that the system has settled to its stable state where in each season one new flu strain appears. We refer to this sequence of strains as the a -lineage. Then in year $T + 1$ two strains appear. Both strains have exactly the same cross-reaction with strains from the previous years while the immunological distance between the two strains corresponds to the distance between strains that are two years apart; this represents a situation where the two strains have arisen through independent mutations from the strain of year T . As

discussed in the introduction, the two strains arrive sequentially in the population such that the epidemic caused by the first strain has already come to a conclusion when the second strain is introduced.

In addition to the permanent immunity responsible for influenza drift, influenza may give rise to a temporary immunity. Ferguson et al. (2003) and Tria et al. (2005) both report that their large simulation models could not produce flu-like phylogenies unless they included a temporary immunity lasting some months and they quote several clinical studies in support of the existence of such short-lived protection. Recently, Forsberg and Christiansen (2003) have found evidence for selection against known TCL-epitopes in human influenza, suggesting that T-cell immunity plays a role. This system may be associated with the observed broad but temporary immunity. Here we will assume that only a fraction v of those infected are available for further infection during the same season, corresponding to a temporary immunity lasting 3–4 months—i.e. a bit longer than suggested by Ferguson et al. (2003).

The strain that is first introduced in the population can be considered to be the next generation of the a -lineage, while the strain that is introduced later will be referred to as the b -strain. Now imagine that strain a has already made an epidemic and that strain b comes to the population. To characterize the immunity to strain b we can no longer rely on identifying the strength of cross-immunity with time since last infection. Instead, the immunity to strain b is determined by summarizing the immunity structure of the hosts after the a -epidemic but prior to the b -epidemic and we subdivide the host population according to the possible immune states

- x_k the fraction of hosts whose last infection occurred k seasons ago and who were not infected by a during season $T + 1$.
- y_k the fraction of hosts who were last infected k seasons ago and who were also infected by a during season $T + 1$.

The size of the immune classes x_k and y_k can be expressed in terms of the immune structure s_i at the onset of the $T + 1$ st epidemic season and ϕ the fraction of hosts that escaped the a -infection in the $T + 1$ st season. Omitting the reference to the season, this yields

$$x_k = \phi s_k, \quad y_k = (1 - \phi) s_k.$$

Since only the immune response to the most related strain causes cross-immunity, the infectivity-reduction factors against b are

Immunity class	Infectivity reduction
x_k	τ_k
y_k for $k > 1$	τ_2
y_1	τ_1

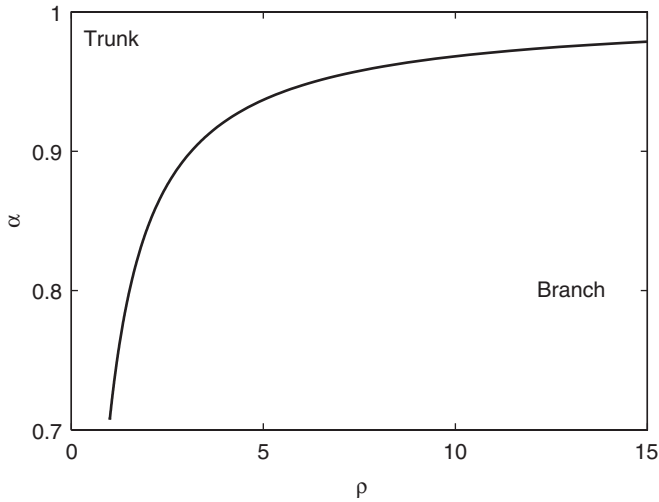


Fig. 3. The smallest width of cross-immunity α that prohibits the establishment of the b -strain and inhibits branching of the phylogenetic tree in the absence of temporary immunity ($v = 1$) as a function of the basic reproduction number ρ .

Now strain b can invade if the threshold condition

$$1 < \rho p_1,$$

holds where p_1 is the potential infectivity of strain b at the onset of the b -epidemic i.e.

$$\begin{aligned} p_1 &= \sum_{k=1}^{\infty} \tau_k x_k + v \tau_1 y_1 + v \sum_{k=2}^{\infty} \tau_k y_k \\ &= \phi q + v \tau_1 (1 - \phi) s_1 + v \tau_2 (1 - \phi) (1 - s_1) \\ &= \phi q + v (1 - \alpha) (1 - \phi) (1 + \alpha \phi). \end{aligned} \tag{7}$$

Here ϕ denotes the fraction of host that did not get infected by the a -strain in the $T + 1$ st season and the last equality holds only when the drift process is at equilibrium.

Figs. 3 and 4 show the region in parameter space that allows the b -strain to cause an epidemic. Clearly the b -strain can cause an epidemic more easily if the disease in question has a high reproduction number than if the reproduction number is low, or more precisely for high ρ the b -strain can establish even if only a fraction of the hosts can be reinfected within a season.

4. The season after the bifurcation

If strain b can produce an epidemic in season $T + 1$, then the b -epidemic will affect the immune structure of the host population including the conditions that will meet the a -lineage in season $T + 2$. In this section we study the transmission dynamics in the subsequent season. Two strains may arise in season $T + 2$, namely the drift progeny of strain a and that of strain b . We shall refer to these two new strains as a_2 and b_2 , respectively, and their immediate ancestors (from season $T + 1$) as a_1 and b_1 . To keep the argument simple we assume that both lineages will produce drift variants that differ by the same amount as that

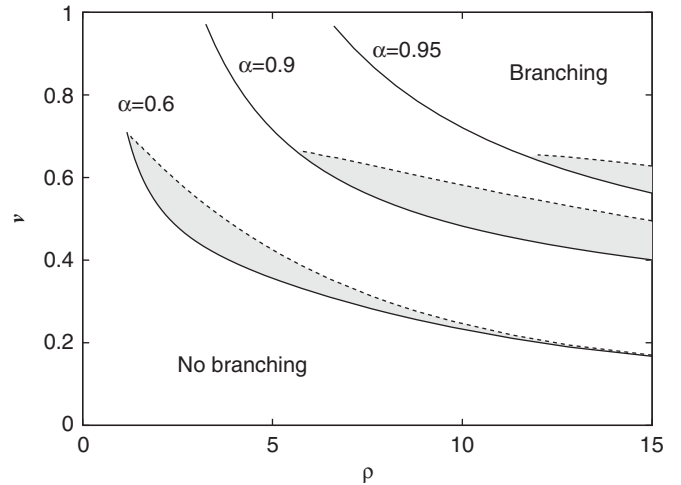


Fig. 4. The outcome of introducing second viral lineage originating from the same ancestor as the epidemic strain. Establishment depends on v the fraction of the host population that can be infected twice in the same season, the reproduction number ρ , and the width of the cross-immune protection α . In the second year after the introduction of a new lineage the outcome may depend on the sequence in which the two strains arrive. Above the full curve strain b can establish during the first season after the mutation. In the shaded area between the full and the broken lines the b strain will eliminate the a -strain if it arrives to the population earlier than the a strain in the subsequent season, $T + 2$. If the a strain arrives first, the b -strain can still establish.

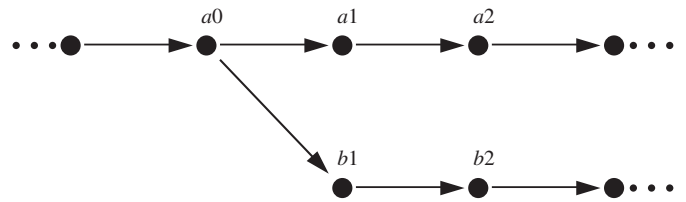


Fig. 5. Relationship between the strains. An arrow goes from ancestor to drift-offspring. Cross-immunity is assumed to depend on the number of steps between the two strains in question.

observed during the normal drift process. Similar to our discussion in the previous section we assume that cross-immunity to the strains is determined by the number of mutations (or drift events) by which the immunizing and challenging strains differ, see Fig. 5.

To describe the cross-immunity structure at the onset of season $T + 2$, we need to determine the size of the epidemics in season $T + 1$. Since the a_1 -epidemic occurred prior to the arrival of the b_1 -strain, the size of the a_1 -epidemic is the equilibrium size $(1 - \phi)$ of the drift epidemics given by Eq. (9). The potential infectivity p_1 of the b_1 -strain is given by Eq. (7) and ψ the fraction of those hosts that could but did not get infected by the b_1 strain is therefore given by

$$\log \psi + \rho p_1 (1 - \psi) = 0.$$

The resulting cross-immunity structure is now determined in Table 1.

Table 1
Immunity classes present at the onset of season $T + 2$ when in season $T + 1$ strains $a1$ and $b1$ have both produced an epidemic affecting the fraction $1 - \phi$, respectively, $1 - \psi$ of the available host population

Symbol	Infection in season $T + 1$		Size of class at onset of $T + 2$	Immunity $a2$	Immunity $b2$
$s_k(ab)$	$a1$	$b1$	$v(1 - \psi)(1 - \phi)s_k$	τ_1	τ_1
$s_k(a0)$	$a1$	0	$(v\psi + (1 - v))(1 - \phi)s_k$	τ_1	$\min\{\tau_3, \tau_{k+1}\}$
$s_k(0b)$	0	$b1$	$(1 - \psi)\phi s_k$	$\min\{\tau_3, \tau_{k+1}\}$	τ_1
$s_k(00)$	0	0	$\psi\phi s_k$	τ_{k+1}	τ_{k+1}

Here s_k is the fraction of the host population whose last infection prior to season $T + 1$ occurred k season earlier.

Two distinct scenarios may occur in season $T + 2$ depending on which of the drift-strains $a2$ and $b2$ that arrive first.

If the $a2$ -strain is the first virus that enters the population it will give rise to an epidemic provided that the threshold condition

$$1 < \rho q_2 = \rho \sum_{k=1}^{\infty} \tau_1 s_k(ab) + \tau_1 s_k(a0) + \min\{\tau_3, \tau_{k+1}\} s_k(0b) + \tau_{k+1} s_k(00). \tag{8}$$

holds. Our numerical investigations of the threshold condition suggest that the threshold-condition is satisfied for all parameter values that allow the $b1$ -strain to produce an epidemic in the previous season. It appears that the cross-immunity produced by the $b1$ -strain is in fact so insignificant that it does not affect the outcome of the $a2$ -epidemic. If the $b2$ -strain subsequently arrives to the population it will meet transmission conditions that are better than those experienced by the $b1$ -strain: the $a2$ -epidemic is slightly smaller and the cross-protection conferred by $a2$ -infections towards the $b2$ -strain is less effective than $a1$ versus $b1$ since $a2$ and $b2$ differ by four steps while $a1$ and $b1$ differ only by two (Fig. 5). Thus, the $b2$ -strain will establish and the splitting of the flu quasi-species into two distinct lineages is completed.

The effect of early arrival of the b -strain in season $T + 2$ could differ significantly from the previous scenario. Since the $b1$ -epidemic is small due to cross-immunity and temporary protection conferred by the $a1$ -epidemic, most of the population will not have significant cross-protection against the $b2$ -strain and consequently this strain may produce a large epidemic, which in turn may prevent an $a2$ -epidemic later in the season.

To determine the size $(1 - \psi_2)$ of the $b2$ epidemic, we first determine p_2 the potential infectivity of the strain. Consulting Table 1, we find that

$$p_2 = \sum_{k=1}^{\infty} \tau_1 s_k(ab) + \min\{\tau_3, \tau_{k+1}\} s_k(a0) + \tau_1 s_k(0b) + \tau_{k+1} s_k(00),$$

such that the fraction of hosts that escape infection in the $b2$ epidemic is the solution to

$$\log \psi_2 + \rho p_2 (1 - \psi_2) = 0.$$

The $b2$ -epidemic affects the epidemic threshold for a subsequent $a2$ -epidemic in two ways since it protects the fraction $1 - v$ of those infected from further infection throughout the season and in addition it confers extra cross-immunity to hosts who had not been infected for at least four seasons. Numerical investigations indicate that the latter effect is insignificant and taking into account only the suppression of $a2$ that is caused by the temporary immunity we find the approximate threshold condition

$$1 < \rho q_2 (1 - \psi_2 + (1 - v)\psi_2) = \rho q_2 (1 - v\psi_2),$$

where q_2 is the potential infectivity of strain $a2$ if it had arrived prior to the $b2$ -epidemic.

The shaded region of Fig. 4 shows the parameters for which the first generation of the b -lineage can invade and its progeny can suppress the a -lineage if it is the first strain to arrive in the subsequent season.

5. Discussion

Figs. 3 and 4 summarize the conditions under which cross-immunity can prevent the establishment of a second drift-mutant within the same epidemic season. For long lasting cross-protection ($\alpha \approx 1$) and small reproduction number ρ , permanent cross-immunity provides sufficient herd immunity to exclude a second invader. For more realistic durations of cross-protection and larger values of ρ , cross-immunity alone cannot prevent branching events and suppression of the second mutant occurs only if a fraction $1 - v$ of those infected obtain a temporary general immunity prohibiting reinfection by other strains within the same epidemic season. Since the empirical support for the existence of a general immune protection is rather weak, we expect that its effect must be small, perhaps on the order of $1 - v \approx 5 - 20\%$ suggesting that the parameter values for influenza must lie in the upper left-hand corner of Fig. 4. As ρ increases and α decreases a higher level of temporary immunity is required to exclude branching.

Since ρ denotes the reproduction number in an immunologically naive population, the quantity cannot be observed directly during drift periods. Estimates based on the first pandemic after a shift suggests that $\rho \approx 2 - 4$ for all three subtypes of influenza A (Mills et al., 2004; Spicer and Lawrence, 1984). For such values of ρ moderate

levels of permanent and temporary cross-immunity can explain the lack of branches in the flu tree. However, the first viral strains arising after an antigenic shift may have uncharacteristically low transmissibility. Bailey's (1986) analysis of the 1965 epidemic in Leningrad gives $\rho \approx 7.5$ with 75% of the population unavailable for infection for an influenza strain occurring more than a decade after the last shift. Thus, a higher level of cross-immunity may be necessary to explain the regular virus drift process.

Hay et al. (2001) recently reviewed the observed evolution of the three influenza drift-lines that circulate in the human population: the Hong Kong subtype, A/H3N2; the Russian subtype, A/H1N1; and the B type. The A/H3N2 influenza exhibits the most slender tree and at any given time little genetic and antigenic variation is observed among the strains collected worldwide. The Russian A/H1N1 and the B influenza both have more branched trees with cocirculation of antigenically distinct strains and both influenzae seems to have split into two successful lineages: B-influenza branched about 25 years ago into the Victoria and the Yamagata lines and the A/H1N1 influenza branched in the mid-90s into the Bayern and the Beijing lines.

It is unclear if differences among influenza (sub-)types in the reproduction number ρ or the width of the cross-immunity α can explain these observations. However, the spatial distribution of the two influenza B lines (and to a lesser extent also that of the lines of A/H1N1) suggests that geographical isolation may be involved in the branching events as well. Thus, the ability of the virus to move successfully between continents may be crucial for the branching process.

While the assumption of homogeneous mixing seems to be critical for our conclusions, we have found that a number of other assumptions may be altered without significantly affecting our findings. Branching conditions are not changed qualitatively if for example cross-immunity acts by reducing susceptibility to reinfection rather than the ability to spread the disease, see Fig. 6. For abrupt decays of cross-immunity the model becomes considerably more complicated and the drift process may lead to regular or irregular oscillations in the annual disease prevalence (Andreasen, 2003). Still our numerical investigations show that our qualitative conclusion holds, with the modification that the invasion threshold may vary over time making branching likely in some seasons and unlikely in other seasons. Similar observations apply to simulations where ρ is varied among seasons to resemble the jumps in cross-immunity over time that has been observed (Smith et al., 2004).

If cross-immunity is almost but not quite strong enough to prevent the establishment of a second mutant and general temporary protection is present, then the mutant strain will cause a minor epidemic in its first season rendering the host population highly susceptible to its progeny in the subsequent season. In fact, if the invader's progeny starts its epidemic before the resident lineage, then

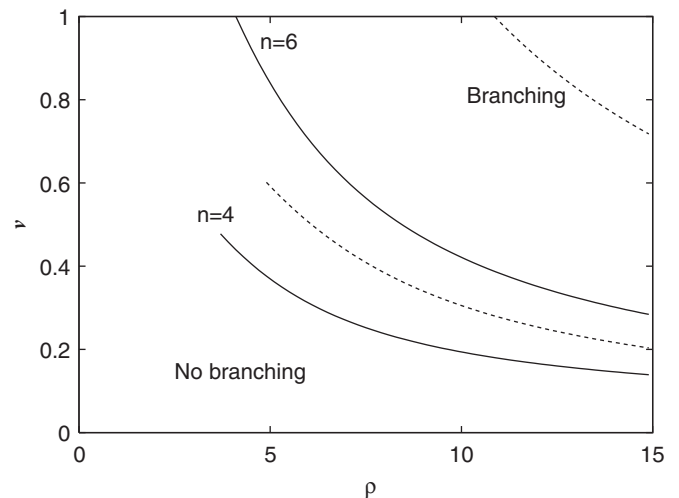


Fig. 6. The outcome of introducing a second viral lineage originating from the same ancestor as in Fig. 4 but for the case where cross-immunity acts by reducing susceptibility with a factor $\sigma_k = (k/n)^m$ for $k < n$ while $\sigma_k = 1$ for $k \geq n$. The factor m controls how significantly cross-immunity acts during the first 1–2 seasons after infection. Strong cross-immunity for a few seasons significantly reduces the chance of branching $m = 2$ (full line), $m = 3$ broken line. For small ρ and small width of cross-immunity ($n = 4$) the drift-equilibrium is unstable and the branching condition varies among seasons (not shown).

the mutant type may give rise to a large epidemic so that the general temporary immunity caused by this epidemic excludes the off-spring from the resident lineage resulting in what would be observed as a large jump in drift type. This phenomenon could explain the occurrence of so-called “herald waves” where next seasons main strain appears in low frequency at the end of the previous season (Glezen et al., 1982) suggesting that herald waves should be associated with large jumps in immunity type and should occur most frequently in influenzae with branched trees.

Recent studies of the population genetics of influenza based on the whole genome show that influenza is considerably more polymorphic than previously thought in particular in its internal structures and that reassortment may generate at least some of the novel viral variants (Holmes et al., 2005). These findings question our interpretation of the relatedness among strains as solely characterized by amino-acid differences in the HA-gene but not the relatednesses themselves as they may be seen as a schematic representation of the observed cross-reactivity.

Our description of influenza drift describes the hypothesis that viable drift variants appear at a high rate and that immuno-selection subsequently weed-out all but a single successful strain. The alternative explanation is that drift is a mutation or recombination limited process where the resident strain would circulate until an immunologically deviating mutant or recombinant establishes; the new strain would then suppress the previous strain. In this scenario the lack of branches would reflect how rare such viable strains are. While this hypothesis certainly deserves more attention, we feel that it is a less likely explanation of drift selection because the immuno-suppression caused by

the new type appears to be rather small compared to the “self-suppression” caused by the epidemic of the original strain itself. To quantify the hypothesis of mutation limitation one would need a better (possible stochastic) description of strain elimination than the one presented here.

Acknowledgment

This research was supported in part by a grant from the National Institutes of Health (NIH) award 1 RO1 GM607929 to S.A. Levin and by Grant 51–00–0392 from the Danish Natural Science Research Council. We thank Martin Nowak and the Institute for Advanced Study for their hospitality during the first stages of this project.

Appendix. Existence and stability of the internal drift equilibrium

The season-to-season dynamics of the drift process is determined by the map

$$F(q) = 1 - \alpha + \alpha\phi q,$$

where $\phi, 0 < \phi \leq 1$ is the solution to Eq. (5). First note that if $q = F(q)$ is an equilibrium value, then the corresponding ϕ and q must solve the equilibrium condition as well as Eq. (5). After elimination of q , we find that ϕ must satisfy the equation

$$G(\phi) = \log \phi + \rho \frac{1 - \alpha}{1 - \alpha\phi} (1 - \phi) = 0, \tag{9}$$

and conversely if ϕ solves the equation $G(\phi) = 0$ then ϕ corresponds to an equilibrium.

Using the method of Andreassen (2003) we will now show that G has a unique zero in the interval $(0, 1)$ if $\rho > 1$ and none if $\rho < 1$. To determine the roots of $G(\phi) = 0$, Andreassen (2003) observes that G has the following properties:

- (1) $G(1) = 0$;
- (2) $G(0^+) = -\infty$;
- (3) $G(\phi)/(1 - \phi)$ is increasing on the interval $(0, 1)$;
- (4) $G'(1) = 1 - \rho$.

If $\rho > 1$, properties (1) and (4) shows that G is positive to the left of $\phi = 1$, and from property (2) we conclude that G has a zero in the interval $(0, 1)$. Property (3) now shows that this zero must be unique. If $\rho < 1$, the function G is negative to the left of $\phi = 1$ and consequently it must have an even number of zeroes on the interval $(0, 1)$ and by property (3) we conclude that there are no zeroes. Since ϕ uniquely determines the equilibrium, we conclude that there exists a unique internal equilibrium when $\rho > 1$ and none if $\rho < 1$.

The stability of the internal equilibrium q is determined by the magnitude of the Jacobian

$$DF(q) = \alpha\phi + \alpha q\phi',$$

where ϕ and ϕ' are evaluated at the equilibrium. By implicit differentiation of Eq. (5) we find that

$$\phi' = \frac{\rho(1 - \phi)}{\rho q - 1/\phi}.$$

Using (5) to eliminate ρq gives

$$\begin{aligned} DF(q) &= \alpha\phi + \alpha \frac{\log \phi}{\log \phi / (1 - \phi) + 1/\phi} \\ &= -\alpha \frac{\phi - 1 - \log \phi}{1/\phi - 1 + \log \phi}. \end{aligned}$$

The elementary inequality $x - 1 - \log x > 0$ applied to numerator and denominator shows that the fraction is positive such that $DF(q) < 0$. Finally, we observe that the difference between the numerator and the denominator is

$$f(\phi) = (\phi - 1 - \log \phi) - (\phi^{-1} - 1 + \log \phi).$$

Since $f(1) = 0$ and $f'(\phi) = (1 - \phi^{-1})^2 > 0$ we conclude that $f < 0$ on the interval $(0, 1)$ such that the fraction numerically is less than unity, showing that $0 > DF(q) > -\alpha > -1$. We conclude that the internal equilibrium is always stable when it exists.

The observations in this appendix simplify considerably the subsequent analysis. The fact that Eq. (9) establishes a one-to-one correspondence between $\rho > 1$ and ϕ the fraction of host that escape infection at equilibrium allows us to use ϕ rather than ρ as our basic parameter thus obtaining explicit analytic expressions for most of the curves we show.

Fig. 3 showing the minimal duration of cross-immunity that can inhibit branching of the tree in the absence of temporary cross-protection, is obtained by setting $p_1 = 1/\rho$ and $v = 1$ in Eq. (7), and combining with the equilibrium condition

$$q = 1 - \alpha + \alpha\phi q \tag{10}$$

plus the definition of ϕ in Eq. (5). After some algebra one finds that

$$\alpha = \frac{1}{\phi} \sqrt{\frac{1}{\log \phi} + \frac{1}{1 - \phi}},$$

and

$$\rho = \frac{(1 - \alpha\phi)(-\log \phi)}{(1 - \alpha)(1 - \phi)}.$$

The curve in Fig. 3 is now parametrized by $\phi, 0 < \phi < 1$. In particular, notice that the smallest α that can inhibit branching for any ρ is given by

$$\lim_{\phi \rightarrow 1} \alpha(\phi) = 1/\sqrt{2}.$$

Similarly the condition for invasion of the b -strain shown in Fig. 4 can be determined as a parametric curve of the form $(\rho(\phi), v(\phi))$, where $\rho(\phi)$ is given above while $v(\phi)$ is found by setting the left-hand side of inequality (7) equal $1/\rho$, solving for v , and expressing q and ρ in terms of ϕ using Eq. (10).

References

- Abu-Raddad, L.J., Ferguson, N.M., 2004. The impact of cross-immunity, mutation and stochastic extinction on pathogen diversity. *Proc. R. Soc. London* 271, 2431–2438.
- Abu-Raddad, L.J., Ferguson, N.M., 2005. Characterizing the symmetric equilibrium of multi-strain host-pathogen systems in the presence of cross immunity. *J. Math. Biol.* 50, 531–558.
- Andreasen, V., 2003. Dynamics of annual influenza A epidemics with immuno-selection. *J. Math. Biol.* 46, 504–536.
- Andreasen, V., Levin, S.A., Lin, J., 1996. A model of influenza A drift evolution. *Z. Angew. Math. Mech.* 76 (Suppl. 2), 421–424.
- Bailey, N.T.J., 1986. Macro-modeling and prediction of epidemic spread at community level. *Math. Modelling* 7, 689–717.
- Boni, M.F., Gog, J.R., Andreasen, V., Christiansen, F.B., 2004. Influenza drift and epidemic size: the race between generating and escaping immunity. *Theor. Popul. Biol.* 65, 179–191.
- Boni, M.F., Gog, J.R., Andreasen, V., Feldman, M.W., 2006. Epidemic dynamics and antigenic drift in a single season of influenza A. *Proc. Roy. Soc. Lond. Ser. B* 273, 1307–1316.
- Buonagurio, D., Nakada, S., Parvin, J., Krystal, M., Palese, P., Fitch, W., 1986. Evolution of human influenza A virus over 50 years: rapid, uniform rate of change in NS gene. *Science* 232, 980–982.
- Cox, N.J., Subbarao, K., 2000. Global epidemiology of influenza: past and present. *Annu. Rev. Med.* 51, 407–421.
- Eigen, M., 1993. Viral quasiespecies. *Sci. Am.* 1993 (7), 32–39.
- Eigen, M., Schuster, P., 1979. *The Hypercycle: A Principle of Natural Selforganization*. Springer, Berlin, Heidelberg, New York.
- Ferguson, N., Andreasen, V., 2002. The influence of different forms of cross-protective immunity on pathogen strain structure. In: Blower, S., Castillo-Chavez, C., Kirschner, D., van den Driessche, P., Yakubu, A. (Eds.), *Mathematical Approaches for Emerging and Reemerging Infectious Diseases: Models, Methods and Theory*. IMA Volumes Mathematics and its Applications, vol. 125. Springer, New York, pp. 157–169.
- Ferguson, N.M., Galvani, A.P., Bush, R.M., 2003. Ecological and immunological determinants of influenza evolution. *Nature (London)* 422, 428–433.
- 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.A., 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.
- Forsberg, R., Christiansen, F.B., 2003. A codon-based model of host-specific selection in parasites, with an application to the influenza A virus. *Mol. Biol. Evol.* 20, 1252–1259.
- Frey, T.K., Abernathy, E.S., Bosma, T.J., et al., 1998. Molecular analysis of rubella virus epidemiology across three continents, North America, Europe, and Asia, 1961–1997. *J. Infect. Dis.* 178, 642–650.
- Girvan, M., Callaway, D.S., Newman, M.E.J., Strogatz, S.H., 2002. Simple model of epidemics with pathogen mutation. *Phys. Rev. E* 65, 031915.
- Glezen, W.P., Couch, R.B., Six, H.R., 1982. The influenza herald wave. *Am. J. Epidemiol.* 116, 589–598.
- Gog, J., Grenfell, B., 2002. Dynamics and selection of many-strain pathogens. *Proc. Natl. Acad. Sci. USA* 99, 17209–17214.
- Gog, J.R., Rimmelzwaan, G.F., Osterhaus, A.D.M.E., Grenfell, B.T., 2003. Population dynamics of rapid fixation in cytotoxic T lymphocyte escape mutants of influenza A. *Proc. Natl. Acad. Sci. USA* 100, 11143–11147.
- Grenfell, B.T., Pybus, O.G., Gog, J.R., et al., 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 303, 327–332.
- Hay, A., Gregory, V., Douglas, A., Lin, Y., 2001. The evolution of human influenza viruses. *Philos. Trans. R. Soc. London B* 356, 1861–1870.
- Holmes, E.C., Ghedin, E., Miller, N., et al., 2005. Whole-genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. *PLoS Biol.* 3 (9), e300.
- Inaba, H., 2001. Kermack and McKendrick revisited: the variable susceptibility model for infectious diseases. *Japan J. Ind. Appl. Math.* 18, 273–292.
- Kawaoka, Y., Gorman, O.T., Ito, T., et al., 1998. Influence of host species on the evolution of the nonstructural (NS) gene of influenza A. *Virus Res.* 55, 143–156.
- Kermack, W.O., McKendrick, A.G., 1927. Contributions to the Mathematical Theory of Epidemics, 1. (Reprinted in *Bull. Math. Biol.* 53, 33–55 (1997)). *Proc. R. Soc.* 115A, 700–721.
- 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.
- Lin, J., Andreasen, V., Casagrandi, R., Levin, S.A., 2003. Traveling waves in a model of influenza A drift. *J. Theor. Biol.* 222, 437–445.
- Mills, C.E., Robins, J.M., Lipsitch, M., 2004. Transmissibility of 1918 pandemic influenza. *Nature (London)* 432, 904–906.
- Ohtsuki, A., Sasaki, A., 2006. Epidemiology and disease-control under gene-for-gene plant pathogen interaction. *J. Theor. Biol.* 238, 780–794.
- 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., Tyrrell, D.A.J., Dickson, 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.
- Rambaut, A., Robertson, D.L., Pybus, O.G., Peeters, M., Holmes, E., 2001. Human immunodeficiency virus—phylogeny and the origin of HIV-1. *Nature (London)* 410, 1047–1048.
- Schweiger, B., Zadow, I., Heckler, R., 2002. Antigenic drift and variability of influenza viruses. *Med. Microbiol. Immunol.* 191, 133–138.
- Shih, S.R., Chen, G.W., Yang, C., et al., 2005. Laboratory-based surveillance and molecular epidemiology of influenza virus in Taiwan. *J. Clin. Microbiol.* 43, 1651–1661.
- Smith, C.B., Cox, N.J., Subbarao, K., Taber, L.H., Glezen, W.P., 2002. Molecular epidemiology of influenza A H3N2 virus infections. *J. Infect. Dis.* 185, 980–985.
- Smith, D.J., Lapedes, A.S., de Jong, J.C., et al., 2004. Mapping the antigenic and genetic evolution of influenza virus. *Science* 305, 371–376.
- Spicer, C.C., Lawrence, C.J., 1984. Epidemic influenza in Greater London. *J. Hyg.-Cambridge* 93, 105–112.
- Thieme, H.R., Yang, J.L., 2002. An endemic model with variable reinfection rate and applications to influenza. *Math. Biosci.* 180, 207–235.
- Tria, F., Lassig, M., Peliti, L., Franz, S., 2005. A minimal stochastic model for influenza evolution. *J. Stat. Mech. Theor. Exp.* P07008.