Ecological and immunological determinants of influenza evolution

Supplementary Information

Neil M. Ferguson[1], Alison P. Galvani[2] & Robin M. Bush[3]

[1]*Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College of Science, Technology and Medicine, St Mary's Campus, Norfolk Place, London W2 1PG, UK*

[2] *Department of Integrative Biology, University of California, Berkeley, CA 94720-3140, USA*

[3] *Dept. of Ecology and Evolutionary Biology, University of California, Irvine, 92697, USA*

Refined immunological model

The metric given in the main text for the antigenic distance between a strain and the immune history of a host can be simply generalised. If each strain is represented by AC amino-acids, $\{j_1(s),..,j_k(s), ..,j_{AC}(s)\}$ (where $1 \leq j \leq 20$), and the immune history of a host is represented by a set of indicator variables $H=\{I[j,k]\}$ where $I[j,k]=0$ if the host has previously been infected with a strain with amino acid j at codon k, and 1 otherwise. Then a more sophisticated model for the distance between a strain s, and a host immune history, H, is given by:

$$d(s,H) = A^{1-1/\gamma}\left[\sum_{n=1}^{A}\left(\sum_{m=1}^{C} I[j_{nC+m}(s), nC+m]\right)^{\gamma}\right]^{1/\gamma}$$

When $\gamma=1$, this reduces to the simpler metric given in the main text, namely the number of codons in strain s for which the amino-acid has not been previously encountered by the host. However, for $\gamma>1$, antigenic distance is greater for multiple changes at a single epitope than for the same number of changes distributed across multiple epitopes, and the reverse holds for $\gamma<1$. For all values of γ, $d(s,H)$ varies between 0 and AC. Long term surveillance of influenza suggests that greater antigenic change results from replacements at multiple epitopes than multiple changes at single epitopes [1,2], suggesting $\gamma<1$. However, from the perspective of the evolutionary statistics examined in the current study, while γ had some effect on the simulated temporal pattern of substitutions, little effect on overall dynamics was observed.

Sensitivity analysis for key parameters

We test the sensitivity of model behaviour to key parameters in order to investigate whether realistic drift dynamics are generated for other parameter choices.

Figure SI1a first illustrates that the model, while intrinsically stochastic, produces realistic patterns of viral evolution for multiple model runs with the same parameters. Figures SI1b-f use time averages of pairwise nucleotide diversity, weekly infection incidence and fixation rate to characterise model behaviour. We see that short-lived non-specific immune protection against reinfection is crucial to reducing equilibrium diversity of the viral population (Figures SI1c-e). For a wide range of other parameters (e.g. mutation rate, cross-immunity intensity), if the duration of short-lived immunity falls below 6 months, viral diversity and infection incidence rapidly reach unrealistic levels. This pattern was seen even for very high levels of cross-immunity between strains with up to four codon differences (Figure SI1h).

The cross-immunity phenotype of the long-lived immune memory response is also a critical determinant of fixation rates (Figure SI1b,d). Three parameters characterise this response: the maximal level of cross-immunity, the threshold number of amino-acid changes beyond which cross-immunity drops below the maximal level, and the minimum level of cross-immunity for strains from the same subtype which share no alleles (at the modelled epitopes) with previously experienced strains (see Methods).. As maximal cross-immunity increases, so do mean fixation rates (Figure SI1b), while diversity and incidence decrease. For a fixed level of maximal cross-immunity, lowering the minimal cross-immunity maximises the selective advantage of novel variants and thus increases fixation rates and lowers diversity (Figure SI1b). Varying mutation rate alone has little effect on diversity or incidence, but decreasing the mutation rate lowers fixation rates (Figure SI1c). Increasing the intensity of cross-immunity by increasing the amino-acid change threshold lowers diversity and incidence (Figure SI1d), though both still reach unrealistically high levels in the absence of a short-lived immune response. The novel inclusion of boosting of the non-specific immune response on exposure to a previously encountered viral strain that fails to result in obvious infection (see Methods) also acts as an

important density-dependent constrain, reducing diversity, incidence and fixation rates (Figure SI1e).

We also tested model sensitivity to the assumption that the short-lived response reduces susceptibility to reinfection. Reducing infectiousness on reinfection (rather than susceptibility) was found to give very similar overall dynamics.

The basic reproduction number, R_0 (ref. 3), and the level of seasonal forcing, ε, were found to have only subtle effects on evolutionary dynamics for reasonable parameter values.

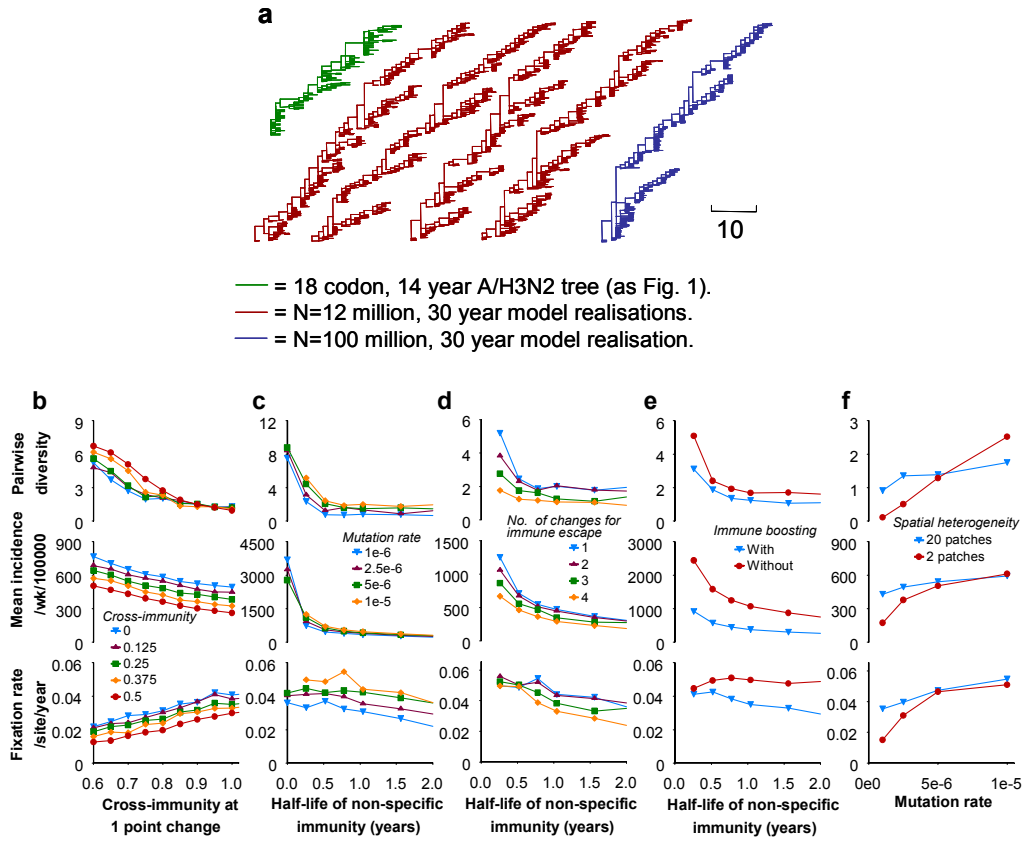

Figure SI1. Sensitivity analyses. *a*. 18 positively selected codon tree for H3 (as Figure 1*d*) compared with trees from 4 realisations of the model with baseline parameters (as Figure 2), and 1 realisation of the model with N=100 million hosts each with lifespan L=60 years in 20 patches arranged in a cylindrical grid with nearest-neighbour patch coupling only, and a mutation rate $\delta=5\times10^{-6}$ *b*. Sensitivity of mean prevalence pairwise genetic distance between coexisting strains, weekly infection incidence and fixation rate to cross-immunity parameters θ_1 (x-axis) and θ_0 (series). Results are averages over 5 or more realisations of 50 years. Parameters: half-life of non-specific immunity $\tau_{1/2}$=1 year, n_t=1, $\delta=2.5\times10^{-6}$. *c*. Sensitivity to changes in $\tau_{1/2}$ and δ. Parameters: n_t=1. *d*. Sensitivity to changes in $\tau_{1/2}$ and n_t. *e*. Sensitivity to whether the non-specific immune response is reset to post-infection levels after exposure events not leading to infection. Parameters: n_t=1, $\delta=2.5\times10^{-6}$. *f*. Sensitivity to spatial heterogeneity (number of geographical patches) and δ. Parameters: $\tau_{1/2}$=1 year, n_t=1, θ_0=0. Unless otherwise stated parameters have baseline values (see Methods). For b-f, results shown are averages of between 5 and 20 model runs.

The results shown assume the host population to be distributed over 20 weakly-coupled geographical patches. This heterogeneity in host population contact patterns was seen to stabilize strain diversity, incidence and fixation rates (Figure SI1f) by reducing the sensitivity of these outcome measures to variation in other parameters (such as mutation rate). This is a result of the enhanced persistence and more stable global dynamics afforded by spatial heterogeneity. The importance of such spatial effects in explaining observed evolutionary patterns can be expected to increase once it becomes feasible to model spread within the entire human population.

Given that it is prohibitively computationally intensive to simulate viral evolution with host population sizes equal to the global human population size, in producing the results above we modelled a population of 12 million and enhanced disease persistence by reducing host lifespan to 30 years (hence increasing birth rates). Critically, these parameter choices ensure that the population size is above the critical population size [4] for single strain persistence. Simulations in smaller host populations can generate influenza-like dynamics across a wide range of parameter space due to unrealistically high single-strain extinction rates arising from the high level of stochastic fluctuation associated with small populations. However, reproducing high observed fixation rates in a population of 12 million requires artificially high mutation rates, resulting in mean incidences of 37% per year – 3 times that observed [5]. It is therefore critical to demonstrate robustness of model dynamics to scaling of host population size. We verified this robustness, comparing results obtained with 12 million hosts with those from a simulated population of 100 million hosts of 60 year lifespan (Figure SI1a). Only the mutation rate and epidemiological coupling between geographical patches needed to be reduced to reproduce realistic evolutionary dynamics (and a reduced 27% annual incidence) at the larger population size. This gives confidence that, when age-related contact rates and increased spatial heterogeneity are included, in the future this model framework will give realistic evolutionary and incidence patterns when applied to populations over 1 billion.

In this context it should be noted that we assumed 12 antigenic codons (rather than 18) for computational reasons: the memory requirements of the model are considerable (2GB of RAM for 12 million hosts, and 12GB for 100 million), and scale linearly with the number of codons modeled. We verified that system dynamics were near identical for 12 and 18 codons (given appropriate minor rescaling of the mutation rate per codon) – an unexpected result given 12 codons already gives a very high-dimensional antigenic space within which evolution can occur. We chose to model 12 codons to enable populations greater than 10 million to be simulated in a reasonable time.

Comparison of Influenza A/H1, A/H3 and B evolution

The sensitivity analyses above give some insight into the potential causes of differences between HI, H3 and B. The lower fixation rate of H1 and B could merely reflect a lower net mutation rate (Figure SI1c), though whether this might be due to a lower point substitution rate or a result of those viruses lower total incidence in the host population is unclear. However, if the diversity (in terms of numbers of co-circulating genetically divergent lineages) of H1 and B is significantly greater than H3 (something it will not be statistically possible to verify without larger numbers of isolates collected over a longer time period), this may indicate the viruses differ with respect to cross-immunity phenotype; *i.e.* H1 and B may have a reduced maximal cross-immunity (Figure SI1b), or a lower threshold number of changes for immune escape (Figure SI1d), giving results similar to those seen in Figure 1h in the main paper. However, a fuller understanding of differences between human influenza viruses must await analysis of the

key loci involved in antigenic selection in H1 and B, and development of a more sophisticated model of subtype coevolution.

Functional constraints on nucleotide substitutions

We extended the model to examine explicitly the effect of functional constraints on viral evolution in three ways:

(a) assuming every new (non-synonymous) variant has a random fitness (i.e. transmissibility) drawn from a uniform distribution of defined width. This model can be thought of a crude representation of changes occurring in parts of the viral genome not being explicitly modelled. Figure SI2(a) shows how increasing the extent of this random variability (by increasing the width of the uniform distribution) has only limited effects on equilibrium diversity and is not able to substitute for short-lived non-specific immunity in explaining observed low levels of diversity.

(b) explicitly modelling a set of codons that determine viral transmissibility independently from those codons determining antigenicity. Unsurprisingly, this model tends to evolve towards higher transmissibility genotypes at those codons. However this evolution occurs over a long timescale and is subject to substantial random fluctuations, since short-term evolution is still dominated by the much more intense competitive selection occurring at the antigenic codons. We found that antigenically novel variants with a 50% transmissibility deficit compared with currently dominant strains were still able to fixate in the population due to their temporary much larger antigenic fitness advantage. Figure SI2(b) shows the limited effect such codons have on viral diversity.

(c) assuming antigenicity and transmissibility are related, such that there is a transmissibility gradient associated with antigenic space, with a very small number of genotypes of the 12 modelled codons having an maximal associated transmissibility. Figure SI2(c) shows that varying the slope of this gradient has little or no effect on the overall pattern of evolution seen, due to the much more intense (albeit frequency-dependent) selective advantage endowed by antigenic changes at those codons.

Clearly, it would be possible to construct a nearly infinite set of such functional constraint models, but the 3 listed above span a reasonable range of feasible possibilities, and all indicate that functional constraints alone are insufficient to explain observed patterns of influenza evolution.

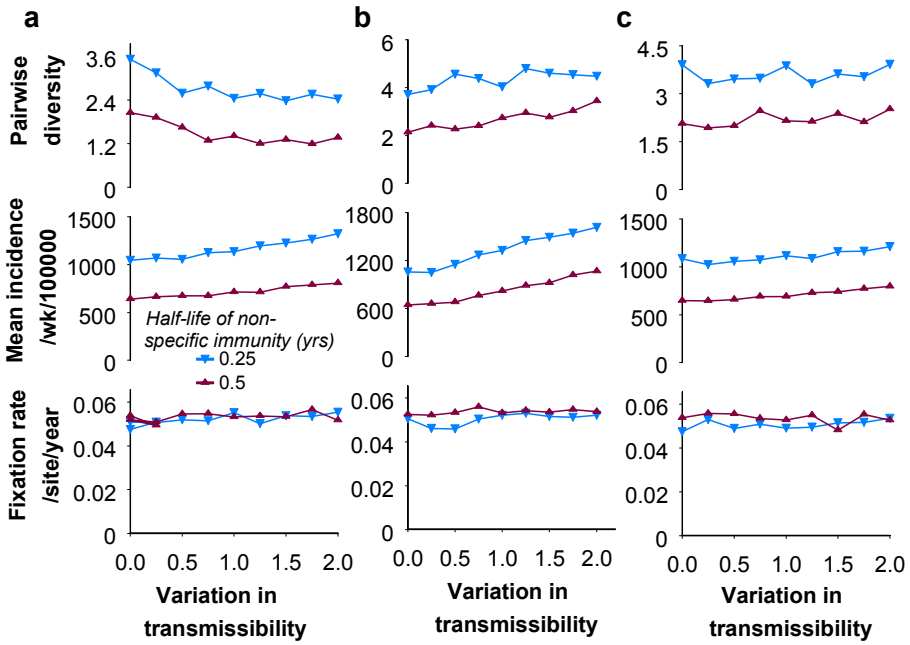

Figure SI2. Functional constraints restricting likely nucleotide substitutions. The three statistics used for Figure SI1 are shown as a function of allowable variation in transmissibility (*i.e.* epidemiological fitness), Δ, defined as Δ=(Max transmissibility- Min transmissibility)/(Median transmissibility). *a.* Model in which each new strain generated via mutation (whether synonymous or non-synonymous) is assigned random transmissibility from a uniform distribution with minimum 1-Δ/2, median 1, and maximum 1+Δ/2. The effect of such unlinked variability on reducing antigenic diversity is noticeable, particularly if transient non-specific immunity is absent or very short-lived, but insufficient to reproduce observed levels of influenza diversity. *b*. Model in which 4 codons (independent of the 12 antigenic codons) determine transmissibility. Every amino-acid is assumed to endow a different transmissibility level, with these levels by dividing the interval [1-Δ/2,1+Δ/2] into 20 discrete values and uniquely assigning these values randomly to the 20 amino acids at the start of each simulation run. The particular mapping of amino acids to transmissibility levels does not effect the results shown. Overall transmissibility of a strain is then determined by averaging across the 4 fitness-determining codons. This functional constraint model results in a fitness gradient within the space of amino acid sequences, with a single sequence having the maximum fitness. The magnitude of the fitness gradient (as determined by Δ) is seen to have no significant effect on diversity or fixation rate. *c.* As b, but the fitness-determining codons are assumed to be the same as the antigenic codons; *i.e.* transmissibility is determined from the amino acid sequence of the 12 antigenic codons in the same manner as used for b, with no independent fitness-determining codons being modelled. This functional constraint model results in the intrinsic fitness of a strain being confounded with antigenically-defined frequency-dependent fitness. Again, the magnitude of the fitness gradient (as determined byΔ) is seen to have no significant effect on diversity or fixation rate. Results shown are averages of between 5 and 20 model runs.

Theoretical argument for the necessity of non-specific transient immunity

From a theoretical perspective, the key issue in understanding influenza evolution is whether lineage diversity can be constrained by strain-specific factors (*i.e.* processes not relying on inter-strain interaction), or whether some form of competitive interaction is essential. In this context, a lineage represents a group of strains that are antigenically similar to each other and distinct from other such sets; *i.e.* strains in a lineage interact at most minimally with other lineages via long lived cross-immunity.

Lineage dynamics are fundamental here since cross-immunity which decays with genetic distance invariably selects for increasing antigenic diversity. Thus the evolutionary optimal (self-organised) strain configuration is a set of non antigenically non-overlapping strains. In the absence of any interaction between lineages, and for a constant host population size and structure, the number of lineages n is governed by a simple birth-death process: $\mathrm{d}n/\mathrm{d}t=\mathbf{b}n - \mathbf{d}n$, where **b** represents the rate at which new lineages emerge through mutation of existing lineages, and **d** the per-lineage rate of extinction.

A number of conclusions arise from considering this simplified model. Firstly, for the system to be in equilibrium, there needs to be a remarkable coincidence of equal values of **b** and **d.** Of course, the system need not be in equilibrium (founder effects caused by occasional pandemics might be able to explain some of the limited diversity of influenza A), but then it is difficult to explain the similarity in diversity levels and tree shapes between influenza A and B, given B does not undergo pandemic shifts. Founder effects are an implausible explanation for a second reason: the rapid timescale of fixation of antigenic changes in influenza; phylogenetic analyses of influenza evolution indicate the number of amino acid changes that accumulate over the order of a 10 year period would be sufficient to all but escape prior cross-immunity.

If the system is in equilibrium, then simple analysis shows that equilibrium is only neutrally stable and fluctuations of order $\sqrt{n}$ would be expected, together with accompanying drift of diversity levels. Given $n\sim1$, this would imply complete extinction of influenza might be a likely outcome, which seems highly implausible in the light of our current understanding of infectious disease persistence and its dependence on host population size.

Indeed, our understanding of disease persistence makes it difficult to conceive that **d** could take any value that was dynamically distinguishable from zero. Measles, pertussis and other highly communicable respiratory 'childhood' diseases (some of which – such as pertussis – have a transmissibility similar to influenza) persist indefinitely in human populations of at most a few million individuals. Thus there is no evidence to support the notion that a single, antigenically stable strain of influenza A or B could not persist likewise – irrespective of the undoubted complexities of human population structure, spatial distributions and contract patterns.

Relaxing the assumption of independent lineages, and returning to the full system, it is possible (see Sensitivity Analysis section above) to generate (at least transiently) flu-like evolutionary dynamics without between-lineage interactions by simulating host population sizes that are below the critical population size (the population size for which extinction becomes highly probable) for a single strain. However such dynamics are not robust to scaling population size, because they rely on single strains going extinct within a short time period in any circumstance. Overall disease persistence for such systems is only ensured by a high mutation rate. However, such systems fundamentally exhibit the same type of birth-death process dynamics as discussed above, meaning the system behaviour is fundamentally relatively unstable.

Implicit to the above discussion is an assumption that new variants are generated fast enough to make immune escape a likely possibility within a relatively short period of time; *i.e.* the space of evolutionary pathways to high fitness is high-dimensional. This assumption may of course be challenged by invoking either functional constraints (*i.e.* most new variants involve nucleotide changes that substantially lower fitness/transmissibility), or more complex antigenic models that dramatically lower the possible number of evolutionary pathways to high fitness.

A range of published deterministic models assume 1 or 2 dimensional discrete antigenic spaces with nearest-neighbour strain interactions (mediated by cross-immunity). Such models generate travelling waves in strain space of constant diversity. However there is no evidence that such strong antigenic constraints exist in the case of influenza. While a fuller understanding must await completion of integrated analyses of antigenic (HA) and genetic data, the sheer variety of substitutions seen in the 18 positively selected codons of HA1, the high fixation rate of substitutions at those codons, the fact that many such variants are seen to co-circulate, and the fact that the codons affect multiple different aspects of the receptor conformation make an assumption of a very low dimensional antigenic space highly unlikely. Put simply, given any one immune history, it appears improbable that there is one new strain that would have a dramatically higher frequency-dependent fitness advantage over all others. That said, it is to be hoped that ongoing work will answer this question more definitively in the near future.

Functional constraints that affect absolute transmissibility (*i.e.* fitness), rather than antigenicity, can more easily be shown not to be able to explain observed dynamics (see results presented in Sensitivity Analysis section above). Firstly, it is worth noting that our model implicitly assumes strong functional constraints are operating; by only modelling 12 codons out of 357 in HA1, we are implicitly assuming no change in the other 345. In reality only about 50% of codons are rigorously conserved. Of the positively selected codons, clearly changes may affect fitness, but given the rapid rate of substitutions seen at those codons, such effects must by definition be limited by comparison with the (transient) frequency dependent fitness advantage conferred by antigenic change.

Finally, having argued that it is difficult to explain observed patterns of influenza evolution using models in which the only form of strain interaction is long-lived cross-immunity, why does a non-specific, short-lived competitive interaction (such as non-specific immunity) robustly reproduce observed patterns of evolution? In essence, the answer lies in the absolute density-dependent constraint on infection prevalence imposed by such an interaction.

The short-lived and initially almost completely protective nature of the immune response is key, since its short timescale means that the overall population effect of such immunity closely tracks instantaneous infection prevalence through time. Therefore, long-lived but less intense cross-immunity between antigenically distant strains does not have the same effect, because it acts over the timescale of the human lifespan, rather than over the timescale of the infection process, thereby merely acting as a reduction of overall viral fitness.

Returning to the simplified lineage model, density dependent constraints on overall infection prevalence almost directly translate into simple density-dependent constraints on lineage diversity: *e.g.* $dn/dt=\mathbf{b}n - (\mathbf{d}+\mathbf{k}(n-n_0))n$, where **k** represents the density-dependent enhancement of lineage extinction rates induced by lineage-transcending competition. Such a system has a stable equilibrium of $n_0+(b-d)/k$, and (for infinitesimal d), can have a vanishingly small probability of overall disease extinction.

Accession Numbers for sequence data from Gen Bank and the Los Alamos Influenza Sequence Database (http://www.flu.lanl.gov/)

Influenza A subtype H3

GenBank: (AF008656-AF008909), (AF180564-AF180666).

Influenza A subtype H1

GenBank: (AB043487-AB043500), (AF026154-AF026155), AF026158, AF055426, AF131993, (AF386774-AF386775), (AF386779-AF386780), (AY029288-AY029292), (D00406-D00407), D00841, (D13573-D13574), D31949, (L19011-L19028), L19549, L20109, L20111, L20113, L20117, L33480, (L33482-L33493), (L33743-L33753), (L33755-L33756), L33758, L33780, M33748, M38353, M59324, M59328, (X00027-X00028), (X00030-X00031), X17221, X59778, Z54289.

*Los Alamos Influenza Sequence Database***:** ISDN13284, ISDN13305, ISDN13427, ISDN13429, ISDNAU0001, (ISDNAU0007-ISDNAU0009).

Influenza B

GenBank: (AB027392-AB027400), (AB027403-AB027408), AB027495, (AB029617-AB029624), (AB029626-AB029628), (AB029631-AB029632), AB033826, (AB036446-AB036451), (AF050060-AF050065), AF050067, (AF059907-AF059930), (AF059932-AF059939), (AF059942-AF059945), AF059948, (AF059950-AF059953), (AF059955-AF059959), (AF059964-AF059967), (AF059969-AF059976), AF059978, (AF059980-AF059989), (AF059991-AF059992), (AF059995-AF060000), (AF060002-AF060004), (AF060006-AF060009), (AF100347-AF100352), (AF100354-AF100356), AF101071, AF129889, AF129896, AF131990, AF131992, (D38644-D38649), K02713, (L19641-L19643), (L19646-L19647), (L76313-L76322), (L76324-L76331), (L76333-L76334), (L76336-L76337), M18384, M21874, (M22943-M22945), (M36105-M36108), M58413, (M58418-M58426), (M65165-M65177), (M76983-M76984), U70384, (X13551-X13553), X53060, X73421.

*Los Alamos Influenza Sequence Database***:** (ISDNAU1000-ISDNAU1004).

References

1. Cox, N. et al. in *Options for the Control of Influenza II.* (eds. Hannoun, C., Kendal, A. P., Klenk, H. D. & Ruben, F. L.) 223-230 (Elsevier Science B.V., Amsterdam, 1993).
2. Plotkin, J. B., Dushoff, J. & Levin, S. A. Hemagglutinin sequence clusters and the antigenic evolution of influenza A virus. *Proc. Natl. Acad. Sci USA* **99**, 6263-6268 (2002).
3. Anderson, R. M. & May, R. M. *Infectious Diseases of Humans; Dynamics and Control* (Oxford University Press, Oxford, 1991).
4. Galvani, A. P., Coleman, R. M. & Ferguson, N. M. The evolution of sex in parasites. *Proceedings of the Royal Society of London, Series B* (in press).
5. Patriarca, P. A. & Strikas, R. A. Influenza vaccine for healthy adults? *N Engl J Med* **333**, 933-4 (1995).