Revealing the competitive dynamics of influenza viruses through evolutionary cartography

Abstract

Introduction

Seasonal influenza infects between 10% and 20% of the human population every year, causing 250,000 to 500,000 deaths annually [1]. While individuals develop long-lasting immunity to particular influenza strains after infection, antigenic mutations to the influenza virus genome result in proteins that are recognized to a lesser degree by the human immune system, leaving individuals susceptible to future infection. The influenza virus population continually evolves in antigenic phenotype in a process known as antigenic drift. A large proportion of the disease burden of influenza stems from antigenic drift; it is why vaccines remain only transiently effective. A thorough understanding of the process of antigenic drift is essential to our efforts to control mortality and morbidity through the use of a seasonal influenza vaccine.

There are currently three major clades of influenza circulating within the human population: influenza A subtype H3N2, influenza A subtype H1N1 and influenza B. Subtypes refer to the genes, hemagglutinin (H or HA) and neuraminidase (N or NA), that are primarily responsible for the antigenic character of a strain. Currently, seasonal influenza is treated with a trivalent vaccine containing one strain of H3N2, one strain of H1N1 and one strain of influenza B. The World Health Organization (WHO) Global Influenza Surveillance Network issues twice-yearly recommendations on which strains of influenza to use as vaccine strains in 9–12 months time, i.e. a February recommendation for the Northern Hemisphere flu season and an August recommendation for the Southern Hemisphere flu season. These recommendations are provided by a panel of experts after review of the available data.

Mutations to the HA1 region of the hemagglutinin protein are thought to drive the majority of antigenic drift in the influenza virus [2]. Experimental characterization of antigenic phenotype is possible through the hemagglutination inhibition (HI) assay [3], which measures the cross-reactivity of one virus strain to serum raised against another strain. Sera from older strains react poorly with more evolved viruses resulting in new strains having

a transmission advantage over established strains. The results of many HI assays across a multitude of virus strains can be combined to yield a two-dimensional map, quantifying antigenic similarity and distance [4]. The antigenic map of influenza A (H3N2) has shown largely linear movement of the influenza virus population since its introduction in 1968. Evolution of antigenic phenotype appears punctuated with periods of stasis interspersed by periods of more rapid innovation, while genetic evolution appears more continuous [4], suggesting that a relatively small number of genetic changes or combinations of genetic changes may drive changes in antigenic phenotype. The process of antigenic drift results in the rapid turnover of the virus population. Although mutation occurs rapidly, standing genetic diversity is low and phylogenetic analysis shows a characteristically 'spindly' tree with a single predominant trunk lineage and transitory side branches that persist for only 1–5 years [5].

Previously, the antigenic and genetic patterns of influenza evolution have been analyzed essentially in isolation. An antigenic map is constructed from a panel of HI measurements, and a phylogenetic tree is constructed from sequence data. However, the opportunity for a combined approach exists as both the antigenic map and the phylogenetic tree often contain many of the same isolates. Here, we implement a flexible Bayesian approach to jointly analyze the antigenic and genetic dynamics of the influenza virus population. We apply this approach to investigate the dynamics of influenza lineages A/H3N2, A/H1N1, B/Victoria and B/Yamagata.

Results

Bayesian multidimensional scaling

In order to assess patterns of antigenic evolution among influenza strains, we implemented a Bayesian probabilistic analog of multidimensional scaling, referred to here as BMDS. In this model, viruses and sera are given N dimensional locations, thus specifying an 'antigenic map', such that distances between viruses and sera in this space are inversely proportional to cross-reactivity. In the BMDS model, a map distance of one antigenic unit translates to an expectation of a 2-fold drop in HI titer between virus and sera. Maps that produce pairwise distances most congruent with the observed titers will have a high likelihood and will be favored by the BMDS model. We integrate over sources of uncertainty, such as antigenic locations, in a flexible Bayesian fashion.

We began with a BMDS analog of the MDS model used by Smith et al. [4], where viruses and sera are represented as 2D locations and HI expectation is relative to the maximum titer of a particular ferret sera. This model performed well, yielding an average absolute predictive error of 0.66 $\log_2$ HI titers on the 6545 training measurements used to build the model, and an average error of 0.86 on an additional 723 test measurements (Table 2). We find that a model with two dimensional antigenic locations better predicts test data than models with higher (or lower) dimensionality (Table 2), and thus specify a two dimensional model in all subsequent analyses. We extend this model by estimating the strength of overall reactivity of each sera rather than fixing this at the maximum

Table 1. Absolute prediction error of $\log_2$ HI titer for training and test data across models.

MDS	Column effects	Row effects	Training error	Test error
1D	Fixed	None	0.91	1.03
2D	Fixed	None	0.66	0.86
3D	Fixed	None	0.65	0.88
4D	Fixed	None	0.71	0.96
5D	Fixed	None	0.71	1.06
2D	Estimated	None	0.55	0.77
2D	Estimated	Estimated	0.50	0.75

Table 2. Antigenic dynamics of influenza lineages.

	A/H3N2	A/H1N1	B/Vic	B/Yam
Rate of population drift	1.004	0.554	0.406	0.205
Population standard deviation	0.787	0.790	0.917	0.631
Rate of drift on side branches	0.067	0.192	-0.024	0.025
Rate of drift on internal branches	0.172	0.017	0.035	0.009
Rate of drift on trunk	1.011	0.465	0.015	0.417
Coefficient of diffusion on side branches	0.425	0.387	0.461	0.251
Coefficient of diffusion on internal branches	0.290	0.161	0.259	0.154
Coefficient of diffusion on trunk	0.511	0.442	1.098	0.260

titer, and additionally, by estimating the strength of reactivity of each virus in calculating an expectated HI titer. We refer to these estimates as column effects and row effects, respectively. We found that including these effects decreased training error to 0.50 and test error to 0.75 $\log_2$ HI titers.

Antigenic drift across influenza lineages

Through this analysis we find that the antigenic phenotype of influenza A (H3N2) underwent rapid turnover from 1968 to 2011 (Figure 1A). Here, cross-reactive measurements only exist betwen strains sampled at most 11 years apart, leaving only threshold titers, e.g. '<40', in more temporally distant comparisons. Because of the threshold of sensitivity of the HI assay, it's impossible to distinguish a linear trajectory in 2D antigenic space, from a curved trajectory, so long as the curve does not bring antigenic phenotype full circle to have cross-reactive measurements between temporally distant strains. To solve this problem of identifiability, we assumed a weak prior that favors linear movement in the 2D antigenic space, with the slope of the linear relationship and the precision of the relationship incorporated into the Bayesian model (see Methods).

We find that influenza A (H3N2) evolved 1.16 antigenic units per year from 1968 to 2011, with 95% highest posterior density (HPD) interval 1.13–1.20 (Figure 1C). However, the rate of antigenic drift was not constant, with year-to-year movement of mean antigenic phenotype showing an interquartile range of 0.40–1.79. Large jumps in antigenic pheno-

type (Figure 1C) correspond to cluster transitions identified by Smith et al. [4]. Most variation is contained within the first antigenic dimension, but dimension 2 ocassionally shows variation when two antigenically distinct lineages emerge and transiently coexist (Figure 1E), as is the case with the previously identified Beijing 1989 and Beijing 1992 clusters.

We find that other types and subtypes of influenza evolved in antigenic phenotype substantially slower than influenza A (H3N2). Influenza A (H1N1) evolved at an average rate of 0.64 antigenic units per year (HPD 0.59–0.70) since its reemergence in 1977 to 2009. H1N1 influenza shows a similar pattern of punctuated antigenic evolution with occasional larger jumps in phenotype, such as the emergence of the Solomon Islands 2006 cluster. It is unclear whether H1N1 undergoes fewer antigenic transitions of comparable magnitude to H3N2, or whether antigenic transitions in H1N1 are of smaller scale. Both the Victoria and Yamagata lineages of influenza B also evolved relatively slowely, with average rates of 0.45 (HPD 0.40–0.51) and 0.20 (HPD 0.15–0.26) antigenic units per year, respectively. From 2000 to 2011, both the Victoria and Yamagata lineages show slow antigenic movement, but no major antigenic transitions.

Competition between influenza lineages

We investigated the relative rates of antigenic evolution in different influenza lineages from 1999 to 2009, finding a general correspondence between antigenic drift along antigenic dimension 1 and relative incidence between influenza A/H3N2, A/H1N1, B/Victoria and B/Yamagata. From 1999 to 2009, influenza A (H3N2) accounted for the majority of antigenic evolution (51%), while A/H1N1, B/Victoria and B/Yamagata split the remainder, accounting for 17%, 18% and 14%, respectively. Similarly, influenza A (H3N2) was responsible for the majority of incidence (54%) in the USA during this time period. Influenza A (H1N1) accounted for 23% of incidence, B/Victoria accounted for 12% of incidence and B/Yamagata accounted for 10% of incidence. These proportions are significantly similar to one another (randomization test $p = 0.010$).

Discussion

Methods

Genetic and antigenic data

We compiled an antigenic dataset of hemagglutination inhibition (HI) measurements for influenza A (H3N2) by collecting data from previous publications [4, 6–8], NIMR vaccine strain selection reports for 2002 and 2008–2012 [9–15] and the Feb 2011 VRBPAC report [16]. We queried the Influenza Research Database [17] and the EpiFlu Database [18] for HA nucleotide sequences by matching strain names, e.g. A/HongKong/1/1968, and only strains for which sequence was present was retained. If a strain had multiple sequences in the databases we preferentially kept the IRD sequence and preferentially kept the longest

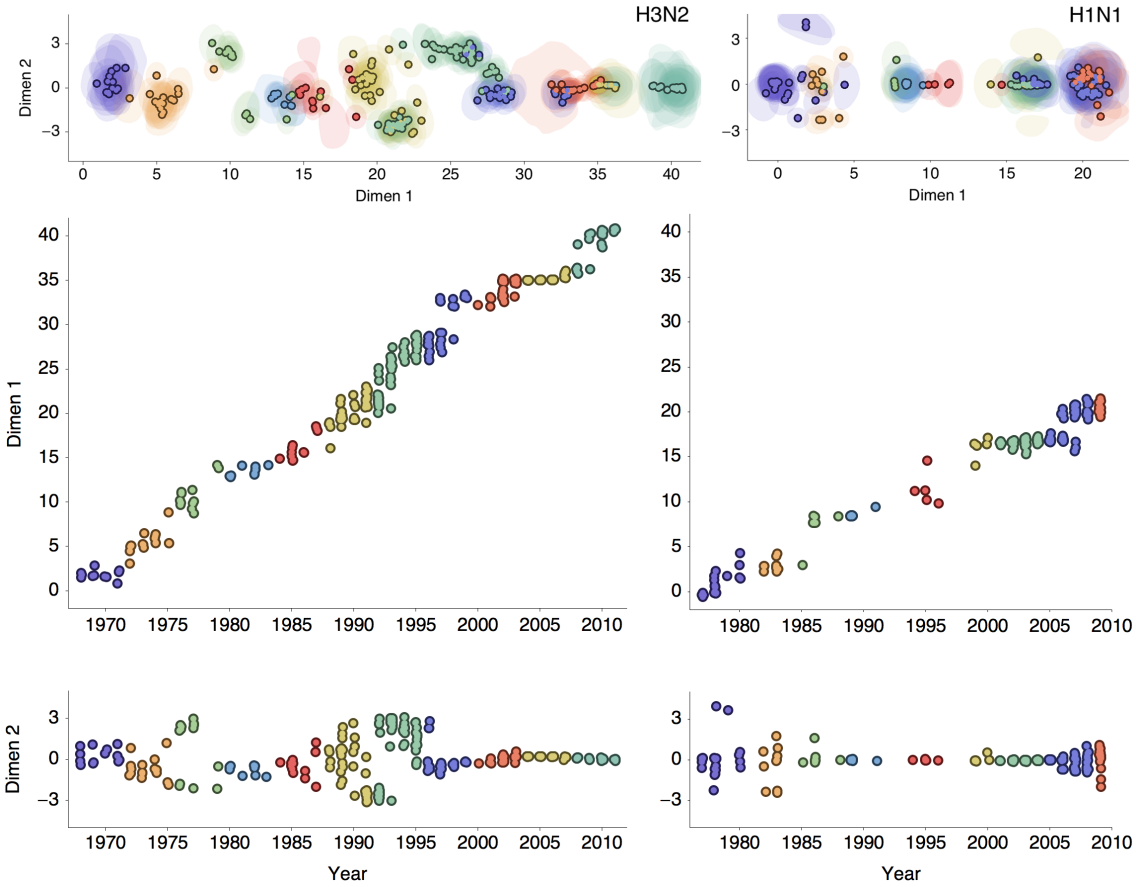

Figure 1. Antigenic locations of influenza H3N2 and H1N1. (A) and (B) Antigenic maps showing the mean posterior location of 338 strains of H3N2 influenza and 243 strains of H1N1 influenza. The model has been constructed so that the primary axis of variation lies along the x-axis (AG1), with the y-axis (AG2) orthogonal to this axis. (C) and (D) Antigenic location along the primary axis of variation (AG1) vs. year of virus isolation. The dashed lines show the relationship of between time and AG1 with a slope of [XXX] for H3N2 and [XXX] for H1N1. (E) and (F) Antigenic location along the secondary axis of variation (AG2) vs. year of virus isolation. Points represent median strain locations and translucent areas represent 50% highest density regions. Antigenic units represent two-fold dilutions of the HI assay, and strains have been colored based on year of isolation.

sequence in IRD. This dataset had 2051 influenza isolates (present as either virus or serum in HI comparisons) dating from 1968 to 2011. However, the majority of isolates were present from 2002 to 2007. Because we are interested in longer-term antigenic evolution, we censored the data to have at most 20 virus isolates per year, preferentially keeping those isolates with more antigenic comparisons. We then kept only those serum isolates that are relatively informative to the antigenic placement of viruses, dropping serum isolates that are compared to 4 or fewer different virus isolates. This censoring left 429 virus isolates, 519 serum isolates and 10,097 HI measurements.

Antigenic data for influenza A (H1N1) was collected from [19–33].

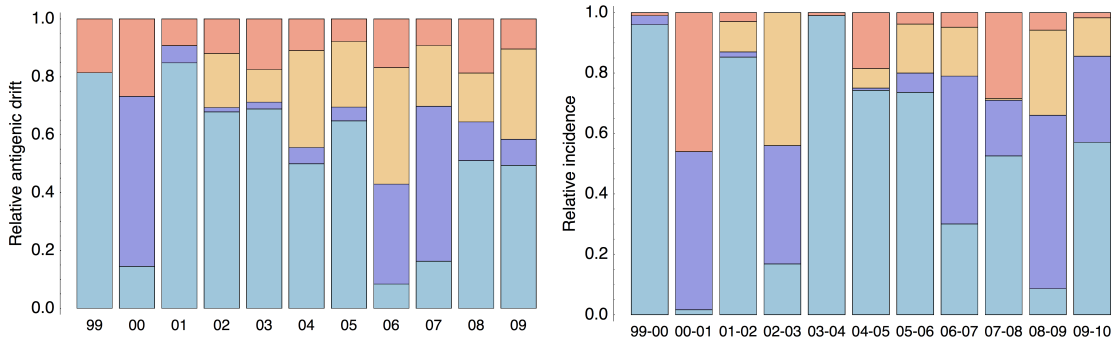

Figure 2. Relative antigenic drift and incidence between lineages of influenza from 1999 to 2009. (A) Relative antigenic drift of influenza A H3N2 and H1N1 and influenza B Victoria and Yamagata lineages. Relative change in antigenic dimension 1 between year i and year $i-1$ is shown for each of the 4 lineages. (B) Relative incidence of influenza A H3N2 and H1N1 and influenza B Victoria and Yamagata lineages. Relative isolation counts of each lineage in the USA for each winter influenza season.

Sequences were aligned using MUSCLE v3.7 under default parameters [34].

Bayesian multidimensional scaling

We follow Smith et al. [4] and represent antigenic locations on a 2D antigenic map. Through the hemagglutination inhibition (HI) assay, there exist measurements of the cross-reactivity of hemagglutinin (HA) from one virus strain to serum raised against another strain [3]. Thus, antigenic phenotype is measured through a series a pairwise comparisons H_{ij}, comparing virus from strain i to serum from strain j. Due to experimental constraints, the distance matrix $\mathbf{H}$ is sparse; most comparisons have not be made. Our goal is to find an optimal projection of the high-dimensional distance matrix into a lower number of dimensions. We conduct this projection using Bayesian multidimensional scaling (BMDS) [35] in which a probabilistic model is constructed to quantify the fit of a particular configuration of cartographic locations to the observed matrix of HI measurements.

Let $X_i \in \Re^P$ represent the cartographic location of virus i for $i = 1, \ldots, n$, and Y_j represent the cartographic location of serum j for $j = 1, \ldots, s$. Virus and antiserum may be isolated from / raised against the same strain and have different cartographic locations. Typically, $P = 2$, but higher or lower dimensions may better reflect the data. This gives set of distances between virus and serum cartographic locations

$$\delta_{ij} = ||X_i - Y_j||_2. \tag{1}$$

Here, H_{ij} represents the $\log_2$ HI titer of virus i against serum j, and HI distance is defined as

$$d_{ij} = \max H_j - H_{ij}. \tag{2}$$

Traditionally, the goal of multidimensional scaling (MDS) optimizes over $\mathbf{X}$ and $\mathbf{Y}$ such that

$$\sum_{(i,j)\in\mathcal{I}} (\delta_{ij} - d_{ij})^2 \tag{3}$$

is minimized, where $\mathcal{I} = \{(\rangle, |) : \mathcal{H}_{\rangle|} \text{ is measured}\}$. Here, we instead assume a probabilistic interpretation in which a point estimate of HI distance is normally distributed

$$d_{ij} \sim \text{Normal}(\delta_{ij}, \sigma^2), \tag{4}$$

and the likelihood of observing some HI distance given the placement of antigenic locations is

$$f_{|}(d_{ij}) = \phi\left(\frac{x - \delta_{ij}}{\sigma}\right), \tag{5}$$

where σ^2 represents variance and $\phi(\cdot)$ represents the standard normal PDF. However, HI assays sometimes show no inhibition of agglutination at all measured titrations, e.g. a measurement can be reported as '<40'. In this case, the likelihood of observing the threshold measurement follows the cumulative density of the upper tail of the normal distribution

$$f_{\lrcorner}(d_{ij}) = 1 - \Phi\left(\frac{x - \delta_{ij}}{\sigma}\right), \tag{6}$$

where $\Phi(\cdot)$ represents the standard normal CDF. Although it is simplest to assume that HI measurements represent point estimates, it seems more natural to assume that the threshold for inhibition occurs between two titers, e.g. we observe inhibition at 1:40 dilution and no inhibition at 1:80. Rather taking the HI titer as 1:40, we can instead treat this as an interval measurement, assuming that the exact HI titer for this measurement would occur somewhere between 1:40 and 1:80. Assuming an HI distance represents an interval estimate, we calculcate its probability according to

$$f_{\sqcup}(d_{ij}) = \Phi\left(\frac{x - \delta_{ij}}{\sigma}\right) - \Phi\left(\frac{x - \delta_{ij} - 1}{\sigma}\right). \tag{7}$$

Throughout the analysis, we use interval probabilities $f_{\sqcup}$ rather than point probabilities $f_{|}$ unless otherwise noted.

We calculcate the likelihood of a set of antigenic locations by multiplying probabilities of individual measurements

$$L(\mathbf{X}) = \prod_{(i,j)\in\mathcal{I}} f(d_{ij}), \tag{8}$$

using probability functions $f_{|}$, $f_{\lrcorner}$ and $f_{\sqcup}$ as appropriate. We assume that the prior precision $\omega = 1/\sigma^2$ follows a diffuse Gamma(a, b) distribution with $a = 0.001$ and $b = 1000$, and we assume a uniform prior over $\mathbf{X}$ and $\mathbf{Y}$.

The preceding model represents HI distance as a drop in titer against the most reactive comparison for a particular antiserum. Here, we relax this assumption and treat the expected homologous titer as a random variable. In this case, HI distance follows

$$d_{ij} = C_j - H_{ij}, \tag{9}$$

with the set of column effects C_j for $j = 1, \ldots, s$ estimated. We assume that C_j values are hierarchically distributed according to a normal distribution. The mean value of this distribution is taken from the empirical mean across columns in the $\log_2$ HI matrix, and the variance across sera is estimated. This formulation assumes that particular sera are

more reactive in general than other sera, we follow the same logic and by including an effective of a particular viruses being more reactive in general than other viruses. Here, HI distance follows

$$d_{ij} = \frac{1}{2}(R_i + C_j) - H_{ij}, \tag{10}$$

and the set of row effects R_i for $i = 1, \ldots, n$ is modeled in an analogous hierarchical fashion, with the variance across viruses estimated. We assume separate Gamma(0.001,1000) hyperpriors on the precisions of the C and R distributions.

We integrate over uncertainty using the Markov chain Monto Carlo (MCMC) procedures implemented in the phylogenetic package BEAST. Metropolis-Hastings proposals include moves to individual virus and serum locations X_i and Y_j, moves to MDS precision ω and moves to C and R precision.

Acknowledgments

Funding

References

1. WHO (2009) Influenza Fact sheet. Available at http://www.who.int/mediacentre/factsheets/fs211/en/.
2. Nelson MI, Holmes EC (2007) The evolution of epidemic influenza. Nat Rev Genet 8: 196–205.
3. Hirst G (1943) Studies of antigenic differences among strains of influenza A by means of red cell agglutination. J Exp Med 78: 407–423.
4. Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, et al. (2004) Mapping the antigenic and genetic evolution of influenza virus. Science 305: 371–376.
5. Fitch WM, Bush RM, Bender CA, Cox NJ (1997) Long term trends in the evolution of H(3) HA1 human influenza type A. Proc Natl Acad Sci USA 94: 7712-8.
6. Hay, AJ and Gregory, V and Douglas, AR and Lin, YP (2001) The evolution of human influenza viruses. Phil Trans R Soc Lond B 356: 1861–1870.
7. Russell CA, Jones TC, Barr IG, Cox NJ, Garten RJ, et al. (2008) The global circulation of seasonal influenza A (H3N2) viruses. Science 320: 340–346.
8. Barr I, McCauley J, Cox N, Daniels R, Engelhardt O, et al. (2010) Epidemiological, antigenic and genetic characteristics of seasonal influenza A (H1N1), A (H3N2) and B influenza viruses: basis for the WHO recommendation on the composition of influenza vaccines for use in the 2009-2010 Northern Hemisphere season. Vaccine 28: 1156–1167.

9. Hay AJ, Lin YP, Gregory V, Bennet M (2002) Annual Report. Technical report, WHO Collaborating Centre for Reference and Research on Influenza, National Institute for Medical Research, UK.

10. Hay A, Daniels RS, Lin YP, Zheng X, Gregory V, et al. (March 2008) Characteristics of human influenza AH1N1, AH3N2, and B viruses isolated September 2007 to February 2008. Technical report, WHO Collaborating Centre for Reference and Research on Influenza, National Institute for Medical Research, UK.

11. Hay A, Daniels RS, Lin YP, Zheng X, Hou T, et al. (Feb 2009) Antigenic and genetic characteristics of human influenza A(H1N1), A(H3N2) and B viruses isolated during October 2008 to February 2009. Technical report, WHO Collaborating Centre for Reference and Research on Influenza, National Institute for Medical Research, UK.

12. McCauley J, Daniels R, Lin YP, Zheng X, Hou T, et al. (Feb 2010) Report prepared for the WHO annual consultation on the composition of influenza vaccine for the Northern Hemisphere. Technical report, WHO Collaborating Centre for Reference and Research on Influenza, National Institute for Medical Research, UK.

13. McCauley J, Daniels R, Lin YP, Zheng X, Hou T, et al. (Sep 2010) Report prepared for the WHO annual consultation on the composition of influenza vaccine for the Southern Hemisphere. Technical report, WHO Collaborating Centre for Reference and Research on Influenza, National Institute for Medical Research, UK.

14. McCauley J, Daniels R, Lin YP, Zheng X, Gregory V, et al. (Sep 2011) Report prepared for the WHO annual consultation on the composition of influenza vaccine for the Southern Hemisphere. Technical report, WHO Collaborating Centre for Reference and Research on Influenza, National Institute for Medical Research, UK.

15. McCauley J, Daniels RS, Lin YP, Zheng X, Gregory V, et al. (Feb 2012) Report prepared for the WHO annual consultation on the composition of influenza vaccine for the Northern Hemisphere. Technical report, WHO Collaborating Centre for Reference and Research on Influenza, National Institute for Medical Research, UK.

16. Cox NJ (Feb 2011) Information for the Vaccines and Related Biological Products Advisory Committee. Seasonal Influenza Vaccines. Technical report, WHO Collaborating Centre for Surveillance, Epidemiology and Control of Influenza, Centers for Disease Control and Prevention, USA.

17. Squires R, Noronha J, Hunt V, García-Sastre A, Macken C, et al. (2012) Influenza research database: an integrated bioinformatics resource for influenza research and surveillance. Influenza and Other Respiratory Viruses DOI: 10.1111/j.1750-2659.2011.00331.x.

18. Bogner P, Capua I, Lipman D, Cox N, et al. (2006) A global initiative on sharing avian flu data. Nature 442: 981–981.

19. Al Faress S, Cartet G, Ferraris O, Norder H, Valette M, et al. (2005) Divergent genetic evolution of hemagglutinin in influenza A H1N1 and A H1N2 subtypes isolated in the south-France since the winter of 2001–2002. Journal of Clinical Virology 33: 230–236.

20. Chakraverty P, Cunningham P, Pereira MS (1982) The return of the historic influenza A H1N1 virus and its impact on the population of the United Kingdom. The Journal of Hygiene 89: 89–100.

21. Chakraverty P, Cunningham P, Shen GZ, Pereira MS (1986) Influenza in the United Kingdom 1982–85. The Journal of Hygiene 97: 347–358.

22. Cox NJ, Bai ZS, Kendal AP (1983) Laboratory-based surveillance of influenza A(H1N1) and A(H3N2) viruses in 1980–81: antigenic and genomic analyses. Bulletin of the World Health Organization 61: 143–152.

23. Daniels RS, Douglas AR, Skehel JJ, Wiley DC (1985) Antigenic and amino acid sequence analyses of influenza viruses of the H1N1 subtype isolated between 1982 and 1984. Bulletin of the World Health Organization 63: 273–277.

24. Daum LT, Canas LC, Smith CB, Klimov A, Huff W, et al. (2002) Genetic and Antigenic Analysis of the First A/New Caledonia/20/99-like H1N1 Influenza Isolates Reported in the Americas. Emerging Infectious Diseases 8: 408–412.

25. Donatelli I, Campitelli L, Ruggieri A, Castrucci MR, Calzoletti L, et al. (1993) Concurrent antigenic analysis of recent epidemic influenza A and B viruses and quantitation of antibodies in population serosurveys in Italy. European Journal of Epidemiology 9: 241–250.

26. Kendal AP, Noble GR, Skehel JJ, Dowdle WR (1978) Antigenic similarity of influenza A(H1N1) viruses from epidemics in 1977–1978 to Scandinavian strains isolated in epidemics of 1950–1951. Virology 89: 632–636.

27. McDonald NJ, Smith CB, Cox NJ (2007) Antigenic drift in the evolution of H1N1 influenza A viruses resulting from deletion of a single amino acid in the haemagglutinin gene. Journal of General Virology 88: 3209–3213.

28. Nakajima K, Nakajima S, Nerome K, Takeuchi Y, Sugiura A, et al. (1979) Genetic relatedness of some 1978–1979 influenza H1N1 strains to 1953 H1N1 strain. Virology 99: 423–426.

29. Nakajima S, Cox NJ, Kendal AP (1981) Antigenic and genomic analyses of influenza A(H1N1) viruses from different regions of the world, February 1978 to March 1980. Infection and Immunity 32: 287–294.

30. Pereira MS, Chakraverty P (1982) Influenza in the United Kingdom 1977–1981. Epidemiology & Infection 88: 501–512.

31. Stevens DJ, Douglas AR, Skehel JJ, Wiley DC (1987) Antigenic and amino acid sequence analysis of the variants of H1N1 influenza virus in 1986. Bulletin of the World Health Organization 65: 177–180.

32. Webster R, Kendal A, Gerhard W (1979) Analysis of antigenic drift in recently isolated influenza A (H1N1) viruses using monoclonal antibody preparations. Virology 96: 258–264.

33. Raymond F, Caton A, Cox N, Kendal A, Brownlee G (1986) The antigenicity and evolution of influenza H1 haemagglutinin, from 1950–1957 and 1977–1983: Two pathways from one gene. Virology 148: 275–287.

34. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797.

35. Oh M, Raftery A (2001) Bayesian multidimensional scaling and choice of dimension. Journal of the American Statistical Association 96: 1031–1044.

Supporting Information