Integrating influenza antigenic dynamics with molecular evolution

Trevor Bedford[1], Marc A. Suchard[2,3,4], Gytis Dudas[1], Philippe Lemey[5], Colin Russell[6], Derek Smith[6,7] & Andrew Rambaut[1,8]

[1]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK
[2]Department of Biomathematics, David Geffen School of Medicine at UCLA, University of California, Los Angeles CA, USA
[3]Department of Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles CA, USA
[4]Department of Biostatistics, UCLA School of Public Health, University of California, Los Angeles CA, USA
[5]Department of Microbiology and Immunology, Katholieke Universiteit Leuven, Leuven, Belgium
[6]Department of Zoology, University of Cambridge, Cambridge, UK.
[7]Department of Virology, Erasmus Medical Centre, Rotterdam, Netherlands.
[8]Fogarty International Center, National Institutes of Health, Bethesda, MD, USA.

April 3, 2012

Abstract

Introduction

[Introduction focuses on vaccine selection to tie things together with the planned conclusion.]

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) and neuraminidase (N), 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 (HA) 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 recent 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, representing antigenic similarity and distance as an easily visualized and quantified measure [4]. The antigenic map of influenza A (H3N2) has shown largely linear movement of the influenza virus population since its introduction in 1968. However, 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 A (H3N2), influenza A (H1N1) and influenza B. [Brief conclusions.]

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, each virus strain is given a N dimensional antigenic location, and the distance between strains in this antigenic space is proportional to their level of cross-reactivity,

as measured by the hemagglutination inhibition (HI) assay. The BMDS model provides an expectation for HI titers, such that a map distance of one antigenic unit between strains translates to a 2-fold drop in HI titer. A map that produces pairwise distances 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 strain location, in flexible Bayesian fashion.

We started with a BMDS implementation of antigenic model used by Smith et al. [4], where each ferret sera and each virus isolate receive a unique location in a 2D antigenic landscape, and map distance is proportional to drop in HI titer relative to the maximum for a particular ferret sera. This model performed well, yielding an average absolute predictive error of 0.71 $\log_2$ HI titers on the 6545 training measurements used to build the model, and an average absolute error of 1.23 on 723 test measurements (Table 1). We extended this model by forcing sera and virus from the same influenza strain to share an antigenic location, reducing the number of required parameters and decreasing test error to 1.05. We further extended this model by sampling the expected titer for the homologous comparison to the sera strain (column effects) and the expected titer for the homologous comparison to the virus strain (row effects). We found that adding these column and row effects, decreased training error to 0.67 and test error to 0.87. We experimented with increasing the number of dimensions beyond two, but found that further increases in dimensionality did not significantly improve test error (Table 1).

Antigenic drift in across influenza types and subtypes

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 enforced a weak prior that favors linear movement in the 2D antigen space. We find that influenza A (H3N2) moved XXX antigenic units per year (XXX–XXX 95% credible interval) from 1968 to 2011 (Figure 1C). However, the rate of antigenic drift was not constant, with year-to-year movement of the mean antigenic phenotype averaging XXX units, but showing an interquartile range of XXX–XXX.

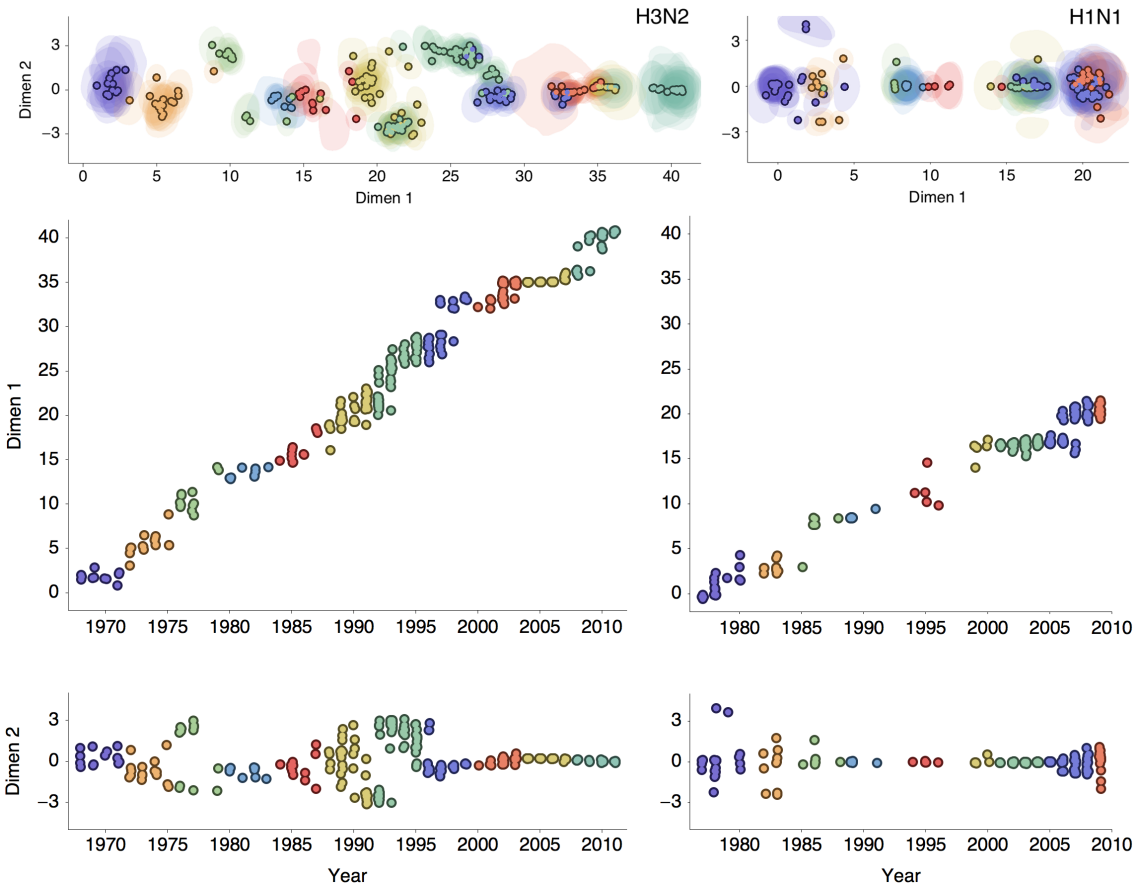

Figure 1. Antigenic locations of influenza H3N2 and H1N1. (A) and (B) Antigenic maps showing the mean posterior location of 338 strains of H3N2 influenza and 243 strains of H1N1 influenza. The map has been oriented so that the primary axis of variation lies along the x-axis (AG1), with the y-axis (AG2) 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. Antigenic units represent two-fold dilutions of the HI assay, and strains have been colored based on year of isolation. [Decide whether to just show mean posterior or show distributions.]

Discussion

Methods

Genetic and antigenic data

We compiled an antigenic dataset for hemagglutination inhibition (HI) measurements for influenza A (H3N2) by combining data used in Hay et al. [6], Smith et al. [4], Russell et al. [7], Barr et al. [8] and Cox et al. [WHO report]. This combined dataset had 1651 influenza isolates (present as either virus or sera in HI comparisons) dating from 1968 to

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

Locations	Column effects	Row effects	MDS	Precision	Training error	Test error
Unique	Fixed	None	2D	0.79	0.71	1.23
Shared	Fixed	None	2D	0.52	0.91	1.20
Shared	Sampled	None	2D	0.62	0.88	1.05
Shared	Sampled	Sampled	2D	0.87	0.72	0.88
Shared	Sampled	Sampled	3D	0.92	0.67	0.87
Shared	Sampled	Sampled	4D	XXX	XXX	XXX
Shared	Sampled	Sampled	5D	XXX	XXX	XXX

2011. However, the majority of isolates date from 2002 to 2007. Because we are interested in longer-term antigenic evolution, we censored the data to have at most 20 strains per year, preferentially keeping those strains with more antigenic comparisons. This censoring left 338 strains present as 320 viruses and 438 sera (replicate sera are often constructed from the same strain). Across these viruses and sera, we observe 7240 HI measurements. We queried the IRD [9] and GISAID [CITE] sequence databases for HA nucleotide sequences based on strain names, e.g. A/HongKong/1/1968, of these strains. If a strain had multiple sequences in the databases we preferentially kept the IRD sequence and preferentially kept the longest sequence in IRD. Sequences were aligned using MUSCLE v3.7 under default parameters [10].

Antigenic data for influenza A (H1N1) was collected from [11–25].

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 sera 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) [26] in which a probabilistic model is constructed to quantify the fit of a particular configuration of cartographic locations to the observed matrix HI measurements.

Let $\mathbf{X}_i \in \Re^P$ represent the cartographic location of strain i for $i = 1, \ldots, N$. Typically, $P = 2$, but higher or lower dimensions may better reflect the data. This gives set of distances between cartographic locations

$$\delta_{ij} = ||\mathbf{X}_i - \mathbf{X}_i||_2. \tag{1}$$

We define the antigenic distance between virus from strain i and antisera from strain j

$$d_{ij} = \max H_{ij} + \log_2 \left(\frac{H_{ij}}{\max H_j} \right). \tag{2}$$

and let the set $\mathcal{I} = \{(\rangle, |) : \mathcal{H}_{)|} \text{ is measured}\}$. The goal of multidimensional scaling (MDS) optimizes over $\mathbf{X}_1, \ldots, \mathbf{X}_N$ such that

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

is minimized. A probabilistic interpretation reformulates the optimization as

$$d_{ij} \sim \text{Normal}(\delta_{ij}, \omega) \times 1(d_{ij} > 0) \text{ for all } (i,j) \in \mathcal{I} \tag{4}$$

where $1(\cdot)$ is the indicator function for truncation. The likelihood of the antigenic distance measures is proportional to

$$\omega^{m/2} \exp\left[-\frac{\omega \times SSR}{2} - \sum_{(i,j)\in\mathcal{I}} \log \Phi(\omega \times \delta_{ij})\right], \tag{5}$$

where m is the cardinality of $\mathcal{I}$, SSR is the sum of the squared residuals and $\Phi(\cdot)$ is the standard normal CDF. We consider the conjugate prior $\omega \sim \text{Gamma}(a, b)$, and assume a uniform prior over $\mathbf{X}$.

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. 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.

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

11. 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.

12. 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.

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

14. 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.

15. 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.

16. 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.

17. 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.

18. 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.

19. 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.

20. 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.

21. 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.

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

23. 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.

24. 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.

25. 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.

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

Supporting Information