Integrating influenza antigenic dynamics with molecular evolution

Trevor Bedford\textsuperscript{1}, Marc A. Suchard\textsuperscript{2,3,4}, Philippe Lemey\textsuperscript{5}, Gytis Dudas\textsuperscript{1}, Victoria Gregory\textsuperscript{6}, Alan J. Hay\textsuperscript{6}, John W. McCauley\textsuperscript{6}, Colin A. Russell\textsuperscript{7,8}, Derek J. Smith\textsuperscript{7,8,9} & Andrew Rambaut\textsuperscript{1,10}

\textsuperscript{1}Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK, \textsuperscript{2}Department of Biomathematics, David Geffen School of Medicine at UCLA, University of California, Los Angeles CA, USA, \textsuperscript{3}Department of Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles CA, USA, \textsuperscript{4}Department of Biostatistics, UCLA Fielding School of Public Health, University of California, Los Angeles CA, USA, \textsuperscript{5}Department of Microbiology and Immunology, Katholieke Universiteit Leuven, Leuven, Belgium, \textsuperscript{6}Division of Virology, MRC National Institute for Medical Research, Mill Hill, London, UK, \textsuperscript{7}Centre for Pathogen Evolution, Department of Zoology, University of Cambridge, Cambridge, UK, \textsuperscript{8}WHO Collaborating Center for Modeling, Evolution, and Control of Emerging Infectious Diseases, University of Cambridge, Cambridge, UK, \textsuperscript{9}Department of Virology, Erasmus Medical Centre, Rotterdam, Netherlands, \textsuperscript{10}Fogarty International Center, National Institutes of Health, Bethesda, MD, USA

Abstract

Influenza viruses undergo continual antigenic evolution allowing mutant viruses to evade host immunity acquired to previous virus strains. Antigenic phenotype is often assessed through pairwise measurement of cross-reactivity between influenza strains using the hemagglutination inhibition (HI) assay. Here, we extend previous approaches to antigenic cartography, and simultaneously characterize antigenic and genetic evolution by modeling the diffusion of antigenic phenotype over a shared virus phylogeny. Using HI data from influenza A subtypes H3N2 and H1N1 and influenza B lineages Victoria and Yamagata, we determine patterns of antigenic drift across lineages and show that antigenic drift drives incidence rates and mediates interference between influenza lineages. Additionally, we describe the selective underpinnings of differences in antigenic drift across lineages. This work makes possible substantial future advances in investigating the dynamics of influenza and other antigenically-variable pathogens by providing a model that intimately combines molecular and antigenic evolution.
**Introduction**

Seasonal influenza infects between 10% and 20% of the human population every year, causing an estimated 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 public health efforts to control mortality and morbidity through the use of a seasonal influenza vaccine.

Before 2009, there were four major clades or lineages of influenza circulating within the human population: the H3N2 and H1N1 subtypes of influenza A, and the Victoria and Yamagata lineages of influenza B. In the case of influenza A, subtypes A/H3N2 and A/H1N1 refer to the genes, hemagglutinin (H or HA) and neuraminidase (N or NA), that are primarily responsible for the antigenic character of a strain. In the case of influenza B, Victoria (B/Vic) and Yamagata (B/Yam) refer to antigenically distinct lineages which diverged from a single lineage prior to 1980 [2]. Mutations to the HA1 region of the hemagglutinin protein are thought to drive the majority of antigenic drift in the influenza virus [3, 4]. Experimental characterization of antigenic phenotype is possible through the hemagglutination inhibition (HI) assay [5], which measures the cross-reactivity of one virus strain to serum raised against another strain through challenge or vaccination. Sera from older strains react poorly with more evolved viruses resulting in new strains having a transmission advantage over previously established strains.

The results of many HI assays across a multitude of virus strains of a single subtype can be combined to yield a two-dimensional map, quantifying antigenic similarity and distance [6]. 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 episodes of more rapid innovation interspersed by periods of stasis, while genetic evolution appears more continuous [6], 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, genetic diversity among contemporaneous viruses 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 [7].

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 characterize jointly the antigenic and genetic evolution of the influenza virus population. We apply this approach to investigate the dynamics of A/H3N2, A/H1N1, B/Vic and
B/Yam viruses, and, for the first time, present detailed reconstructions of the antigenic dynamics of all four circulating influenza clades.

Results and discussion

Antigenic and evolutionary cartography

In order to assess patterns of antigenic evolution among influenza strains, we implemented a Bayesian probabilistic analog of multidimensional scaling (MDS), referred to here as BMDS (see Materials and methods). 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 apply this model to HI measurements of virus isolates against post-infection ferret antisera for influenza A/H3N2, A/H1N1, B/Vic and B/Yam (see Materials and methods).

We begin with Bayesian analogs of the models used by Smith et al. [6], in which viruses and sera are represented as N-dimensional locations, expected titer is relative to the maximum titer of a particular ferret serum and virus and serum locations follow a diffuse prior. After comparing models of differing dimensions, Smith et al. [6] arrive at a 2D model as the preferred model for their data. Smith et al. [6] implement a form of MDS, seeking to optimize virus and serum locations such that the sum of squared errors between expected and observed titers is minimized (eq. 3 of Materials and methods). Here, in implementing BMDS, we provide a likelihood function for the probability of observing HI data given virus and serum locations (eq. 8 of Materials and methods) and seek to estimate model parameters through Bayesian inference using Markov chain Monte Carlo (MCMC). However, the basic antigenic model describing drop in HI titer as proportional to Euclidean distance between virus and serum locations is identical between these methods.

We test model performance by constructing training datasets representing 90% of the HI measurements for each of the four influenza lineages and test datasets representing the remaining 10% of the measurements for each lineage. By fitting the BMDS model to the training dataset, we are able to predict HI titers in the test dataset and compare these predicted titers to observed titers. We find that a two dimensional model has better predictive power than models of lower or higher dimension in all four influenza lineages (models 1–5; Table 1). We find that this 2D model performs well, yielding an average absolute predictive error of between 0.78 and 0.91 log₂ HI titers across influenza lineages (model 2; Table 1), in line with the results of Smith et al. [6]. Consequently, we specify a two dimensional model in all subsequent analyses. The finding of a low-dimensional map across influenza lineages extends previous studies in A/H3N2 [6] and remains an interesting and fundamental empirical observation.
Table 1. Average absolute prediction error of log₂ HI titer for test data across models and datasets.

<table>
<thead>
<tr>
<th>Model</th>
<th>Dimen</th>
<th>Location prior</th>
<th>Serum effects</th>
<th>Virus effects</th>
<th>A/H3N2</th>
<th>A/H1N1</th>
<th>B/Vic</th>
<th>B/Yam</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1D</td>
<td>Uninformed</td>
<td>Fixed</td>
<td>None</td>
<td>1.35</td>
<td>0.94</td>
<td>0.90</td>
<td>1.08</td>
</tr>
<tr>
<td>2</td>
<td>2D</td>
<td>Uninformed</td>
<td>Fixed</td>
<td>None</td>
<td>0.91</td>
<td>0.78</td>
<td>0.82</td>
<td>0.90</td>
</tr>
<tr>
<td>3</td>
<td>3D</td>
<td>Uninformed</td>
<td>Fixed</td>
<td>None</td>
<td>0.93</td>
<td>0.80</td>
<td>0.85</td>
<td>0.92</td>
</tr>
<tr>
<td>4</td>
<td>4D</td>
<td>Uninformed</td>
<td>Fixed</td>
<td>None</td>
<td>0.98</td>
<td>0.84</td>
<td>0.90</td>
<td>0.97</td>
</tr>
<tr>
<td>5</td>
<td>5D</td>
<td>Uninformed</td>
<td>Fixed</td>
<td>None</td>
<td>1.04</td>
<td>0.89</td>
<td>0.98</td>
<td>1.04</td>
</tr>
<tr>
<td>6</td>
<td>2D</td>
<td>Drift</td>
<td>Fixed</td>
<td>None</td>
<td>0.91</td>
<td>0.75</td>
<td>0.77</td>
<td>0.83</td>
</tr>
<tr>
<td>7</td>
<td>2D</td>
<td>Diffusion/Drift</td>
<td>Fixed</td>
<td>None</td>
<td>0.89</td>
<td>0.74</td>
<td>0.74</td>
<td>0.83</td>
</tr>
<tr>
<td>8</td>
<td>2D</td>
<td>Diffusion/Drift</td>
<td>Estimated</td>
<td>None</td>
<td>0.77</td>
<td>0.73</td>
<td>0.66</td>
<td>0.75</td>
</tr>
<tr>
<td>9</td>
<td>2D</td>
<td>Diffusion/Drift</td>
<td>Estimated</td>
<td>Estimated</td>
<td>0.76</td>
<td>0.71</td>
<td>0.64</td>
<td>0.72</td>
</tr>
</tbody>
</table>

Previous work on influenza antigenic and genetic evolution has examined antigenic and genetic patterns in isolation, and then compared the results to assess congruence of these two aspects of evolution [6, 8, 9]. Here, we simultaneously model antigenic and genetic evolution by adopting an evolutionary diffusion process [10], wherein a virus’s antigenic character state evolves along branches of the phylogenetic tree according to a Brownian motion process (see Materials and methods). The phylogenetic diffusion process acts as a prior on virus locations, so that genetically similar viruses are expected to share similar antigenic locations. The antigenic diffusion process includes both systematic drift with time and covariance induced by phylogenetic proximity. We examine the effects of including only systematic drift (model 6; Table 1) and systematic drift plus phylogenetic diffusion (model 7; Table 1), finding a small increase in predictive accuracy of between 0.02 and 0.08 log₂ HI titers when both processes are included. The systematic drift process informs virus and serum locations by dates of isolation and the phylogenetic diffusion process informs virus locations by genetic sequences. Thus, in these models, antigenic locations are inferred using both genetic data and HI data and will differ from locations inferred from HI data alone. If HI data is rich, then we expect minor differences in antigenic locations with the inclusion of genetic data (as may be the case for A/H3N2), while if HI data is spare, then we expect genetic data to play a larger role in determining antigenic locations (as may be the case for B/Vic and B/Yam).

We further extend the model by estimating the strength of overall reactivity of each serum rather than fixing this at its maximum titer, and additionally, by estimating the strength of reactivity of each virus isolate. We refer to these estimates as serum effects and virus effects, respectively. We find that including these effects decreases test error further, ranging in improvement from 0.03 to 0.13 log₂ HI titers (models 8 and 9; Table 1).

We find that the average absolute error in predicted log₂ HI titer is nearly constant with antigenic distance (Figure 1A), thus supporting our model assumption that the drop in log₂ titer is proportional to the Euclidean distance separating viruses and sera on the antigenic map. Additionally, we find that the absolute error in predicted titer is nearly constant with time (Figure 1B). Antigenic locations inferred by the model are well resolved; estimates of antigenic distance between pairs of viruses show relatively little variation across the
posterior. We estimate that virus distances have, on average, a 50% credible interval of ±0.45 antigenic units for A/H3N2, ±0.57 units for A/H1N1, ±0.76 units for B/Vic, and ±0.65 units for B/Yam.

Figure 1. Absolute error in log$_2$ HI titer against antigenic distance and date of virus sampling in influenza A/H3N2. (A) Points show the absolute deviation from the predicted log$_2$ HI titer of the antigenic model, computed from virus location $x_i$, serum location $y_j$, virus effect $v_i$ and serum effect $s_j$, based upon antigenic distance between virus and serum $\delta_{ij}$. The red line shows a LOESS regression curve of these points. (B) Points show the absolute deviation from the predicted log$_2$ HI titers of the antigenic model, based upon the date of virus sampling. The red line shows average deviations for each year. Relationships for A/H1N1, B/Vic and B/Yam are similar.

We find strong correspondence between our results and previous results by Smith et al. [6], with equivalent models producing globally consistent antigenic maps and other models producing locally consistent maps with a small degree of global inconsistency (Figures 8–11). Differences in the MDS and BMDS approaches reflect greater philosophical differences between maximum-likelihood and Bayesian statistical approaches, with the former seeking the single most likely explanation for the data, and the latter seeking to fully characterize model uncertainty. If a single map is desired, then the MDS method may be preferred, while the BMDS method provides flexibility, allowing other sources of information, such as genetic data, to be more easily incorporated.

Antigenic evolution across influenza lineages

Through our analysis, we reveal the antigenic, as well as evolutionary, relationships among viruses in influenza A/H3N2, A/H1N1, B/Vic and B/Yam, quantifying both antigenic and evolutionary distances between strains (Figure 2). Over the time period of 1968 to 2011, influenza A/H3N2 shows substantially more antigenic evolution than is exhibited by A/H1N1 over the course of 1977 to 2009 or B/Vic and B/Yam over the course of 1986 to 2011. We observe prominent antigenic clusters in A/H3N2 and A/H1N1, but less prominent, though still apparent, clustering in B/Vic and B/Yam. Antigenic clusters show high genetic similarity, so that we observe very few mutation events leading to each cluster, rather than the repeated emergence of clusters. This analysis makes the fate of antigenic clusters obvious, with two clusters in A/H3N2 (Victoria/75 and Beijing/89) appearing to be evolutionary dead-ends. Labeling of prominent antigenic clusters in figures 2 and 3
is intended as a rough guide for orientation and not as exhaustive catalog of antigenic variation.

![Figure 2](image)

**Figure 2. Antigenic locations of A/H3N2, A/H1N1, B/Vic and B/Yam viruses showing evolutionary relationships between virus samples.** Circles represent a posterior sample of virus locations and have been shaded based on year of isolation. Antigenic units represent two-fold dilutions of the HI assay. Absolute positioning of lineages, e.g. A/H3N2 and A/H1N1, is arbitrary. Lines represent mean posterior diffusion paths when virus locations are fixed. Prominent antigenic clusters are labeled after vaccine strains present within clusters, and are abbreviated from Hong Kong/68, England/72, Victoria/75, Bangkok/79, Sichuan/87, Beijing/89, Beijing/92, Wuhan/95, Sydney/97, Fujian/02, California/04, Wisconsin/05, Brisbane/07, Perth/09 (A/H3N2), USSR/77, Singapore/86, Beijing/95, New Caledonia/99, Solomon Islands/06 (H1N1), Victoria/87, Hong Kong/01, Malaysia/04, Brisbane/08 (Vic), Yamagata/88, Shanghai/02, Florida/06, Wisconsin/10 (Yam).

HI assays lack sensitivity beyond a certain point, so that for A/H3N2, cross-reactive measurements only exist between strains sampled at most 14 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 difficult to distinguish a linear trajectory in 2D antigenic space from a slightly curved trajectory. To solve this problem of identifiability, we assumed a weak prior that favors linear movement in the 2D antigenic space (present in models 6 through 9; Table 1), with the slope of the linear relationship and the precision of the relationship incorporated into the Bayesian model (see Materials and methods). Because of this, we interpret map locations locally rather than globally, and assess rates of antigenic movement without making strong statements about the larger configuration under which the movement occurs.

We find that influenza A/H3N2 evolved along antigenic dimension 1 at a rate of 1.01 antigenic units per year, with a 95% highest posterior density (HPD) interval of 0.98–1.04 (Figure 3). However, we observe occasional large jumps in antigenic phenotype (Figure 3), corresponding to cluster transitions identified by Smith et al. [6]. Most variation is contained within the first antigenic dimension, but dimension 2 occasionally shows variation when two antigenically distinct lineages emerge and transiently coexist (Figure 2), as is the case with the previously identified Beijing/89 and Beijing/92 clusters.
Figure 3. Antigenic drift of A/H3N2, A/H1N1, B/Vic and B/Yam viruses showing evolutionary relationships between virus samples. Antigenic drift is shown in terms of change of location in the first antigenic dimension through time. Circles represent a posterior sample of virus locations and have been shaded based on year of isolation. Antigenic units represent two-fold dilutions of the HI assay. Relative positioning of lineages, e.g. A/H3N2 and A/H1N1, in the vertical axis is arbitrary. Lines represent mean posterior diffusion paths when virus locations are fixed. Prominent antigenic clusters are labeled after vaccine strains present within clusters, and are abbreviated from Hong Kong/68, England/72, Victoria/75, Bangkok/79, Sichuan/87, Beijing/89, Beijing/92, Wuhan/95, Sydney/97, Fujian/02, California/04, Wisconsin/05, Brisbane/07, Perth/09 (A/H3N2), USSR/77, Singapore/86, Beijing/95, New Caledonia/99, Solomon Islands/06 (H1N1), Victoria/87, Hong Kong/01, Malaysia/04, Brisbane/08 (Vic), Yamagata/88, Shanghai/02, Florida/06, Wisconsin/10 (Yam).

We find that other clades of influenza evolved in antigenic phenotype substantially slower than A/H3N2 (Figure 3). Influenza A/H1N1 evolved at a rate of 0.62 units per year (HPD 0.56–0.67), but showing a similar pattern of punctuated antigenic evolution with occasional larger jumps in phenotype, such as the emergence of the Solomon Islands/06 cluster. Influenza B/Victoria also evolved relatively slowly, with an average rates of 0.42 (HPD 0.32–0.51) units per year. Influenza B/Yamagata evolved slower still, with an
average rate of 0.32 (HPD 0.25–0.39) units per year. Punctuated evolution is less obvious in B/Yam and B/Vic compared to A/H3N2 and A/H1N1, but antigenic clusters are still apparent, with recent transitions to the Brisbane/08 cluster in B/Vic [11] and to the Wisconsin/10 cluster in B/Yam [12]. Interestingly, a minor lineage of B/Vic, denoted B/Hubei-Songzi/51/2008 [11], has persisted through 2011, while remaining antigenically distinct from B/Brisbane/60/2008 viruses (Figure 3).

We sought to summarize year-to-year patterns of antigenic drift by calculating the difference in mean location between consecutive years (Figure 4). We estimate year-to-year antigenic drift for years 1992 to 2011 by calculating the average location along dimension 1 of phylogenetic lineages present in the tree at year \( i \) and comparing this location to the average location of phylogenetic lineages present in the tree at year \( i - 1 \). There may often be large discontinuities in virus locations across the population; our use of difference in mean location is meant to capture both the distance between antigenic clusters and also the change in cluster frequency over consecutive years. We observe greater heterogeneity in year-to-year antigenic drift in type A than in type B lineages (Figure 4), with standard deviation of year-to-year antigenic drift equal to 0.97 units in A/H3N2, 0.66 units in A/H1N1, 0.46 units in B/Vic and 0.26 units in B/Yam. This analysis classifies drift only to the level of consecutive years; some coarse-graining of the timings of transition events will necessarily occur.

**Epidemiological consequences of antigenic drift**

We investigate the relationship between rates of antigenic drift and USA incidence in A/H3N2, A/H1N1, B/Vic and B/Yam. We take estimates of the rate of antigenic drift from the preceding section and compare these rates to measurements of relative incidence for each influenza clade calculated from the proportion of influenza viruses attributable to each clade (see Materials and methods). We analyze incidence from the 1998/1999 to the 2008/2009 seasons to avoid possible complications from the 2009 pandemic, finding that A/H3N2 accounted for 56%, A/H1N1 for 20%, B/Vic for 10% and B/Yam for 14% of viruses detected in the USA during this time period. We find a significant correlation between rate of antigenic drift and relative incidence across the four clades (Pearson correlation, \( r = 0.96, p = 0.021 \)).

We follow-up this analysis with a more detailed analysis of year-to-year variation in antigenic drift and clade-specific incidence in the USA. We measure year-to-year antigenic drift within each clade as described in the previous section and shown in Figure 4. We measure seasonal incidence of each clade by taking average influenza-like illness (ILI) percentage and multiplying this by the proportion of viruses attributable to a clade for each season. This measure of incidence has previously been shown to have have predictive power in the analysis of seasonal influenza trends [13]. In correlating antigenic drift to seasonal incidence we standardize this measure of incidence across seasons and across clades to measure effects in terms of standard deviations of incidence. We find that years with high levels of antigenic drift coming into an influenza season, e.g. drift in A/H3N2 from 2000 to 2001, show high levels of incidence in that season, e.g. incidence of A/H3N2 in the 2001/2002 season (Figure 4, Figure 5A). Here, within-clade drift explains 35% of the variance in stan-
Figure 4. Year-to-year antigenic drift between 1992 and 2009 in A/H3N2, A/H1N1, B/Vic and B/Yam viruses. (A) Time series of year-to-year antigenic drift across influenza clades from 1992 to 2011 and seasonal incidence from 1998/1999 to 2008/2009. Colored lines represent year-to-year antigenic drift, and drift for a virus clade for year $i$ is measured as the mean of antigenic dimension 1 of phylogenetic lineages in year $i$ compared to the mean of antigenic dimension 1 of phylogenetic lineages from the previous year $i-1$. For example, 2000 represents difference in antigenic dimension 1 between 1999 and 2000. Error bars represent 50% Bayesian credible intervals of year-to-year drift. Gray dotted lines represent clade-specific USA incidence taken as average influenza-like illness (ILI) multiplied by proportion of viruses attributable to a clade for each season. Here, 2000 represents the 2000/2001 influenza season. (B) Histograms of year-to-year antigenic drift across influenza clades summarizing observations from 1992 to 2011.

dstandardized incidence and represents a significant correlation (Pearson correlation, $r = 0.60$, $p = 1.6 \times 10^{-5}$).

We further show that incidence within a clade decreases with antigenic drift in other clades, e.g. drift in A/H1N1, B/Vic and B/Yam from 2000 to 2001 decrease incidence of A/H3N2 in the 2001/2002 season (Figure 5B). This also represents a significant relationship (Pearson correlation, $r = -0.20$, $p = 0.024$). We fit a linear model that predicts standardized incidence based upon drift within a clade, drift in other clades of the same virus type, and drift in clades of differing type (Table 2). For example, standardized incidence in A/H3N2 would be explained by drift within A/H3N2, drift in A/H1N1 and by drift across types in B/Vic and B/Yam. We compared adjusted $R^2$ across models with different regression coefficients included, and arrived at a best-fitting model that explains 46% of the variance in standardized incidence and includes the terms $\beta_c$, $\beta_s$ and $\beta_t$, but does not include an intercept term. We find that 1 unit of antigenic drift within a clade increases incidence by
Figure 5. Relationship between antigenic drift within and between clades and seasonal USA incidence for years 1998 to 2009. (A) Standardized incidence within a clade compared to antigenic drift within the same clade. For example, incidence of A/H3N2 in season 1998/1999 is measured against year-to-year antigenic drift of A/H3N2 from 1997 to 1998. (B) Standardized incidence within a clade compared to antigenic drift in other clades. For example, incidence of A/H3N2 in season 1998/1999 is measured against year-to-year antigenic drift of A/H1N1 from 1997 to 1998, as well as antigenic drift of B/Vic from 1997 to 1998. Standardized incidence is measured as yearly ILI percentage multiplied by proportion of viruses identified in a particular clade, and then standardized so that mean and variance across seasons and clades is equal to 0 and 1 respectively. Colors represent the clade in which incidence is measured; light blue for A/H3N2, purple for A/H1N1, orange for B/Vic and red for B/Yam.

1.26 standard deviations, while 1 unit of drift in a clade of the same type decreases incidence by 0.61 standard deviations and 1 unit of drift in clades of differing type decreases incidence by 0.32 standard deviations. The term for across lineage interference $\beta_t$ is small in magnitude, but remains statistically significant (Table 2).

Table 2. Estimates of regression coefficients for the relationship of clade-specific USA incidence to antigenic drift within and across clades, where standardized incidence in A/H3N2 is predicted by $\beta_c x_{H3} + \beta_s x_{H1} + \beta_t x_{Vic} + \beta_t x_{Yam}$ and standardized incidence in other clades follows an analogous relationship.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_c$</td>
<td>1.26</td>
<td>0.22</td>
<td>$1.1 \times 10^{-6}$</td>
</tr>
<tr>
<td>$\beta_s$</td>
<td>-0.61</td>
<td>0.22</td>
<td>0.008</td>
</tr>
<tr>
<td>$\beta_t$</td>
<td>-0.32</td>
<td>0.11</td>
<td>0.005</td>
</tr>
</tbody>
</table>

These findings clearly demonstrate the epidemiological importance of antigenic drift and show that incidence across A/H3N2, A/H1N1, B/Vic and B/Yam is strongly influenced by levels of antigenic drift within each clade, consistent with antigenic drift increasing the susceptible pool available to a clade. The finding of a negative correlation between incidence within one clade and antigenic drift in other clades supports the hypothesis that epidemiological interference exists between clades [13,14]. Interestingly, we find that interference between sister clades, e.g. B/Vic and B/Yam, appears stronger than interference between more divergent clades, e.g. B/Vic and A/H3N2. If interference is mediated
by CTL epitopes, then we would expect interference to decrease with evolutionary distance [15].

Furthermore, these findings suggest that the predominant clade of a particular influenza season could be estimated ahead of time by examining the degree of variation in antigenic drift between influenza clades. However, this prediction would not be perfect, and 56% of the variance in incidence remains unexplained by these estimates of antigenic drift. Some of this variance may be explained by incorporating geographic variation in year-to-year antigenic drift; although there exists significant geographic variation in clade-specific incidence [16], we calculated antigenic drift from global samples, rather than USA-specific samples.

Conversion of antigenic mutation into fixed differences

We observe a faster rate of antigenic drift in influenza A/H3N2 than in A/H1N1 or in either lineage of influenza B. Previous work using general epidemiological models has suggested that rates of antigenic drift may be influenced by both the fundamental reproductive number \( R_0 \) and the rate at which mutation decreases cross-immunity [17, 18]. Owing to these and similar findings, subsequent work ascribed differences in the evolution of A/H3N2 relative to A/H1N1 and influenza B to either greater \( R_0 \) or greater mutation rate [14, 19]. Here, we examine the conversion of recently acquired antigenic mutation into fixed differences by comparing rates of antigenic diffusion to the degree of persistence on phylogeny branches in A/H3N2, A/H1N1, B/Vic and B/Yam (Table 3). To facilitate comparisons, we measure the diffusion coefficient \( D \) [20] characterizing the antigenic movement of each influenza lineage (see Materials and methods). This coefficient is measured in terms of units\(^2\)/year and can be used to determine the diffusion length scale, that is the expected relationship between time and distance traveled, which is \( 2\sqrt{tD} \) for a two-dimensional diffusion. Higher diffusion coefficients mean a lineage is diffusing faster through antigenic space. We compare diffusion coefficients between trunk branches and side branches, where trunk branches are defined as branches ancestral to the most recent virus sample and side branches are defined as those branches not ancestral to the most recent virus sample.

Table 3. Estimates of fundamental diffusion coefficient \( D \) across influenza lineages and phylogenetic branches, including posterior means and 50% highest posterior density intervals.

<table>
<thead>
<tr>
<th>Lineage</th>
<th>Trunk branch ( D )</th>
<th>Side branch ( D )</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/H3N2</td>
<td>0.90 (0.79–0.97)</td>
<td>0.62 (0.57–0.65)</td>
<td>1.45 (1.31–1.57)</td>
</tr>
<tr>
<td>A/H1N1</td>
<td>0.62 (0.50–0.71)</td>
<td>0.44 (0.38–0.48)</td>
<td>1.45 (1.18–1.70)</td>
</tr>
<tr>
<td>B/Vic</td>
<td>0.63 (0.51–0.71)</td>
<td>0.62 (0.54–0.68)</td>
<td>1.02 (0.87–1.14)</td>
</tr>
<tr>
<td>B/Yam</td>
<td>0.35 (0.29–0.40)</td>
<td>0.33 (0.28–0.36)</td>
<td>1.09 (0.93–1.23)</td>
</tr>
</tbody>
</table>

Mutations must proceed through the sieve of natural selection, first appearing as transient differences restricted to side branches, while advantageous mutations will preferentially fix in the virus population and thus accrue on the trunk of the phylogeny. We expect the diffusion coefficient on side branches to more closely resemble the underlying rate of antigenic mutation, before selection can strongly influence observed rates. Here, we find
that A/H3N2 and B/Vic have similar diffusion coefficients on side branches, which are larger than the diffusion coefficients of A/H1N1 and B/Yam (Table 3). Thus, we expect that new antigenic mutations may occur at a faster rate in A/H3N2 and B/Vic than in A/H1N1 or B/Yam. We find that trunk branches in both A/H3N2 and A/H1N1 evolve in antigenic phenotype faster than side branches (Table 3), suggesting that positive selection is promoting antigenic movement in these lineages. Interestingly, B/Vic and B/Yam show less of a signature of positive selection, with similar diffusion coefficients on trunk and side branches (Table 3). Thus, although drift forward along dimension 1 is driven by selection in all lineages (Figure 3), raw antigenic diffusion appears to be selected for more strongly in influenza A than in influenza B, giving influenza A an advantage in overall rate of antigenic drift. Additionally, it appears that A/H3N2 and B/Vic benefit from a greater influx of new antigenic mutations, giving them a similar advantage over their sister lineages A/H1N1 and B/Yam, respectively.

In this context, it is especially interesting that adaptive evolution in the HA1 protein determined through analysis of nonsynonymous and synonymous substitutions is substantially stronger in A/H3N2 than in A/H1N1 [21]; H3 appears to accumulate adaptive fixations at \( \sim 150\% \) the rate of H1 [22]. Here, we observe that A/H3N2 has a diffusion coefficient \( \sim 141\% \) greater than A/H1N1 along side branches of the phylogeny. These results suggest that the greater rate of adaptive fixation in A/H3N2 may be simply due to a greater abundance of new antigenic mutations to fix.

We suggest that the larger signal of positive selection for antigenic diffusion in influenza A relative to influenza B could be due to either a greater fundamental reproductive number \( R_0 \) or to pleiotropic effects associated with antigenic change. In the latter case, we expect that many antigenic variants may be at an inherent disadvantage against other strains from decreased protein function [23, 24] or other effects, even though they benefit from a transmission advantage mediated by host immunity. These mutations may not be promoted by natural selection to the same extent as antigenic variants that maintain functional equivalence. Additionally, other work has shown strong pleiotropic constraints in the evolution of oseltamivir resistance in A/H1N1 [25]. Still, it is clear that more work is necessary to fully clarify the mechanistic underpinnings of the relative evolutionary success of different influenza lineages.

**Conclusions**

In this study we provide a foundation for evolutionary antigenic cartography, which seeks to simultaneously assess antigenic phenotype and antigenic evolution. We use this approach to characterize competitive dynamics across influenza clades A/H3N2, A/H1N1, B/Vic and B/Yam and show that antigenic evolution drives strain replacement in each clade and interference between clades. We find that varying levels of mutational input and varying efficiencies of the conversion of antigenic polymorphism into fixed differences results in different rates of antigenic drift across influenza clades. Influenza A/H3N2 benefits from a higher rate of new antigenic mutation coupled with strong adaptive evolution and evolves in antigenic phenotype faster than A/H1N1, B/Vic and B/Yam. Correspondingly,
we observe substantially greater levels of incidence in A/H3N2 than in other influenza clades. We suggest that antigenic evolution strongly influences competitive dynamics both within and between influenza clades.

The statistical framework presented here represents a baseline to which further advancements in modeling antigenic phenotype and evolution may be made. For example, our likelihood-based model facilitates the inclusion of possible covariates affecting immunological titer, which could include experimental factors such as red blood cell type used in the HI assay [26] and whether oseltamivir is included in the HI reaction [27]. Additionally, this framework should be ideally suited to uncovering genetic determinants of antigenic change, as both the sequence state and antigenic location of internal nodes in the phylogeny may be estimated. In this fashion, it should be possible to correlate sequence substitutions directly to antigenic diffusion. Here, methods to efficiently calculate branch- and site-specific codon substitutions may be necessary to easily correlate sequence substitutions with antigenic change [28].

A deeper understanding of the process of antigenic evolution in the human influenza virus may eventually prove to be crucial in improving vaccine strain selection.

Materials and methods

Antigenic cartography

Antigenic characteristics of viral strains are often assessed through immunological assays such as the hemagglutination inhibition (HI) assay [5]. At heart, these assays compare the reactivity of one virus strain to antibodies raised against another virus strain via challenge or vaccination. In the case of HI, the measurement of cross-reactivity takes the form of a titer representing the dilution factor at which serum raised against a particular virus ceases to be effective at inhibiting the binding of another virus to red blood cells. These factors are commonly assessed by serial dilution, so that HI titers will form a log series, 40, 80, 160, etc. . . .. Because experimental HI titers typically differ by factors of two, we find it convenient to work in logarithm space and represent the titer of virus $i$ against serum $j$ as $H_{ij} = \log_2(\text{HI titer})$, i.e. a titer of 160 has $H_{ij} = 7.32$. Due to experimental constraints, most comparisons cannot be made, leading to a sparse observation matrix $H = \{H_{ij}\}$. Further, measurements are usually interval and truncated, e.g. inhibition may cease somewhere between the serial titers of 160 and 320, or inhibition may be absent at all titers assayed, suggesting a threshold somewhere between 0 and 40.

Previous work [6, 29] has used multidimensional scaling to place viruses and sera on an ‘antigenic map’. These methods heuristically optimize locations of viruses and sera by seeking to minimize the sum of squared errors between titers predicted by map locations and observed titers. Antigenic maps produced by these methods have proved useful in categorizing virus phenotypes [6], but the extension of these methods to integrate genetic data remains notably lacking.

Here, we follow previous models in representing antigenic locations as points in a low $P$-
dimensional antigenic map. One of our initial goals is to find an optimal projection of the high-dimensional distance matrix $H$ into this lower dimensional space. We conduct this projection using Bayesian multidimensional scaling (BMDS) [30] in which we construct a probabilistic model to quantify the fit of a particular configuration of cartographic locations to the observed matrix of serological measurements. Typically, $P = 2$, but higher or lower dimensions may better reflect the data.

Let $x_i \in \mathbb{R}^P$ represent the cartographic location of virus $i$ for $i = 1, \ldots, n$, so that $x_i = (x_{i1}, x_{i2})'$ for $P = 2$. Similarly, let $y_j$ represent the cartographic location of serum $j$ for $j = 1, \ldots, k$, so that $y_j = (y_{j1}, y_{j2})'$ for $P = 2$. For notational compactness, we collect together all virus coordinates into an $n \times P$ matrix $X = (x_1, \ldots, x_n)'$ and all serum coordinates into an $k \times P$ matrix $Y = (y_1, \ldots, y_k)'$. Virus and serum may be isolated from / raised against the same strain and have different cartographic locations, and separate serum isolates raised against the same strain may also have different cartographic locations. This gives a set of distances between virus and serum cartographic locations

$$\delta_{ij} = ||x_i - y_j||_2,$$

where $|| \cdot ||_2$ is an $L_2$ norm.

Traditional approaches to antigenic cartography [6] begin by defining immunological distance as

$$d_{ij} = s_j - H_{ij},$$

where $H_{ij}$ is the log$_2$ titer of virus $i$ against serum $j$ and serum effect $s_j = \max(H_{1j}, \ldots, H_{nj})$ is fixed. In following multidimensional scaling (MDS), these approaches attempt to optimize over unknown $X$ and $Y$ such that

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

is minimized, where $I = \{(i, j) : H_{ij}$ is measured$\}$. In the case of threshold measurements, this error function is modified slightly; see [6] for further details.

Here, we instead assume a probabilistic interpretation in which an observed titer is normally distributed around its cartographic expectation with variance $\varphi^2$,

$$H_{ij} \sim N(s_j - \delta_{ij}, \varphi^2).$$

Consequently, the likelihood of observing an exact titer given the placement of antigenic locations is

$$f_i(H_{ij}) = \phi\left(\frac{H_{ij} + \delta_{ij} - s_j}{\varphi}\right),$$

where $\phi(\cdot)$ represents the standard normal probability density function (PDF). Previous BMDS has employed a sampling density truncated to strictly positive quantities since $d_{ij}$ are directly observed, non-negative quantities. In the antigenic setting, these remain random and can be negative since neither $s_j$ is known nor is $H_{ij}$ observed with much precision.
Here we show the likelihoods of observing three different outcomes given \( \delta_{ij} = 4, \varphi = 0.95 \) and \( s_j = \log_2(1280) = 10.32 \). The likelihood of observing a threshold titer of ‘\(<40\)’ is equal to the lower tail of the probability density function \( f_j(5.32) = 0.146 \). The likelihood of observing a point measurement with an exact inhibiting titer of ‘90.5’ is equal to the density function \( f_j(6.5) = 0.413 \). The likelihood of observing an interval measurement with an inhibiting titer somewhere between ‘160’ and ‘320’ is equal to \( f_s(7.32) = 0.129 \).

HI assays sometimes show no inhibition 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 lower tail of the normal distribution

\[
f_j(H_{ij}) = \Phi\left(\frac{H_{ij} + \delta_{ij} - s_j}{\varphi}\right), \tag{6}
\]

where \( \Phi(\cdot) \) represents the standard normal cumulative distribution function (CDF). Although it is simplest to assume that immunological 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 1:160 dilution and no inhibition at 1:320 dilution. Rather than taking the HI titer as 160, we can instead treat this as an interval measurement, assuming that the exact titer for inhibition would occur somewhere between 160 and 320. HI titers are usually reported as the highest titer that successfully inhibits virus binding, so that in this case, we calculate the likelihood of an interval measurement as

\[
f_s(H_{ij}) = \Phi\left(\frac{H_{ij} + \delta_{ij} - s_j + 1}{\varphi}\right) - \Phi\left(\frac{H_{ij} + \delta_{ij} - s_j}{\varphi}\right). \tag{7}
\]

These likelihoods are illustrated in Figure 6. Throughout our analyses, we use interval likelihoods \( f_s \) rather than point likelihoods \( f \) unless otherwise noted.

We calculate the overall likelihood by multiplying probabilities of individual measurements

\[
L(X, Y) = \prod_{(i, j) \in I} f(H_{ij}), \tag{8}
\]
using probability functions $f_i$, $f_j$, and $f_{ij}$ as appropriate. We begin by assuming independent, diffuse normal priors on virus and serum locations

$$x_i \sim \text{Normal}(\mu, \Sigma)$$
$$y_j \sim \text{Normal}(\mu, \Sigma),$$

where $\mu = (0, \ldots, 0)'$ and $\Sigma$ is a diagonal matrix with diagonal elements all equal to 10000.

**Virus and serum effects**

The preceding model represents immunological distance as a drop in titer against the most reactive comparison for a particular serum. However, this model may be biased in some circumstances. In one example, if a particular serum $j$ is only measured against distant viruses, its maximum titer will be artificially low and the likelihoods concerning this serum will appear poor. To address this issue, we relax the assumption of fixed $s_j$ values and treat the expected log$_2$ titer when $\delta_{ij} = 0$ as a random variable. In this case, $H_{ij}$ still follows equation 4 with expectation $s_j - \delta_{ij}$, but the vector of ‘serum effects’ $\mathbf{s} = (s_1, \ldots, s_k)$ is random and estimated rather than fixed. We assume that $s_j$ values are hierarchically distributed according to a normal distribution. We take an Empirical Bayesian approach in specifying the mean and variance of this distribution, set to the empirical mean and empirical variance of the set of maximum titers across sera $\{\max(H_{1j}, \ldots, H_{nj}) : j = 1, \ldots, k\}$. This formulation assumes that particular sera are more reactive in general than other sera.

Additionally, we follow the same logic and assume that some virus isolates are more reactive than other virus isolates and include a ‘virus effect’ $v_i$ representing the general level of reactivity across HI assays. With virus reactivity included, observed titers follow

$$H_{ij} \sim \mathcal{N}\left(\frac{v_i + s_j}{2} - \delta_{ij}, \sigma^2\right),$$

and the vector of virus effects $v_i$ for $i = 1, \ldots, n$ is estimated in an analogous hierarchical fashion, with $\mathbf{v}$ normally distributed with mean and variance equal to the empirical mean and variance of the set of maximum titers across viruses $\{\max(H_{i1}, \ldots, H_{ik}) : i = 1, \ldots, n\}$.

**Drift model of antigenic evolution**

As presented, multiple configurations of virus and serum locations $\mathbf{X}$ and $\mathbf{Y}$ will give the same likelihood of an observed data matrix $\mathbf{H}$. An example of this phenomenon is shown in Figure 7. In this case, it is impossible to determine from the HI data at hand whether the blue and yellow viruses are antigenically similar (Figure 7A) or antigenically divergent (Figure 7B). This presents an issue of model identifiability, where absolute, as opposed to relative, antigenic locations cannot be determined from the observing the serological data alone. Thus, in order to achieve more a interpretable model we impose a weak prior on global locations. In influenza, it’s clear that antigenic distance between strains increases with time [6, 29]. To capture this, we replace our previous diffuse prior with an informed
Figure 7. Schematic antigenic map with three viruses and two sera. (A) Map with virus 1 and virus 3 antigenically similar. (B) Map with virus 1 and virus 3 antigenically divergent. Virus 1 is shown in blue, virus 2 is shown in red and virus 3 is shown in yellow. Virus isolates are represented by filled circles, sera raised against viruses are shown as open circles and map distances $\delta_{ij}$ are shown as solid lines connecting viruses and sera. Sera from virus 1 is compared against viruses 1 and 2, while sera from virus 2 is compared against viruses 2 and 3. Configurations (A) and (B) represent cartographic models that would give equal likelihoods to a set of serological data $\{H_{11}, H_{21}, H_{22}, H_{32}\}$.

prior in which the expected location of viruses and sera increases with date of sampling along dimension one, and each virus and serum location follows an independent normal distribution centered around this temporal expectation, so that

\[
\begin{align*}
x_{i1} &\sim \beta t_i + \mathcal{N}(0, \sigma^2_x) \\
y_{j1} &\sim \beta t_j + \mathcal{N}(0, \sigma^2_y),
\end{align*}
\]

(11)

where $t$ is the difference between the date of the indexed virus or serum and the date of the earliest sampled virus or serum, and other dimensions follow $x_{im} \sim \mathcal{N}(0, \sigma^2_x)$ and $y_{jm} \sim \mathcal{N}(0, \sigma^2_y)$ for $m \geq 2$. Thus, this model assumes that virus and serum locations drift in a line across the antigenic map at rate $\beta$. The parameter $\sigma_x$ determines the breadth of the cloud of virus locations at each point in time, while $\sigma_y$ determines the breadth of the cloud of serum locations.

Phylogenetic diffusion model of antigenic evolution

We simultaneously model antigenic locations and genetic relatedness by assuming that virus locations are influenced by evolution following a Brownian motion process [10]. To do this, we replace the previous prior specifying independent virus locations with a prior that incorporates covariance based on shared evolutionary history

\[
X \sim \begin{pmatrix} \beta t_1 & 0 \\ \vdots & \vdots \\ \beta t_n & 0 \end{pmatrix} + \text{Evolutionary Brownian Process}(\sigma_x, \tau) \tag{12}
\]

for $P = 2$, where $\sigma_x$ is the volatility parameter of the Brownian motion and $\tau$ is a phylogeny specifying tree topology and branch lengths. Thus, viruses which are genetically
similar are induced to have prior locations close to one another on the antigenic map. In the evolutionary Brownian process, the tips of the phylogeny $\tau$ correspond to the set of virus locations $(x_1, \ldots, x_n)$, and the probability of observing tip locations depends on the locations of internal nodes $(x_{n+1}, \ldots, x_{2n-2})$ and on the location of the root node $x_{2n-1}$. This process assumes that a virus location $x_i$ follows from the location of its parent virus $x_{f(i)}$, and with the addition of drift along dimension 1, is distributed as

$$x_i \sim (\beta d_i, 0)' + N(x_{f(i)}, d_i \Sigma)$$

for $P = 2$, where $f(i)$ is a function that maps nodes to parental nodes, $d_i$ is the length of the branch connecting virus $i$ to parent virus $f(i)$, and $\Sigma$ is a diagonal matrix with diagonal elements all equal to $\sigma^2_x$. The root virus location $x_{2n-1}$ is assumed to follow a normal distribution with expectation $(\beta t_{2n-1}, 0)'$ for $P = 2$ and variance determined by the diffusion volatility $\sigma_x$ [10]. The probability of virus locations $p(X|\beta, \sigma_x, \tau)$ is determined through analytical integration across internal states following the methods introduced in [10]. This formulation corresponds to a Wiener process with drift, in which the drift term $\beta$ only influences the expected states of nodes along the phylogeny, but does not influence the covariance structure among these nodes, which remains the same as it does in a standard Wiener process [31]. This allows the separation in equation 12 between drift terms affecting only expectations and the evolutionary Brownian process that includes covariance among virus locations $x_1, \ldots, x_n$.

Here, the phylogenetic tree $\tau$ is estimated using sequence data for viruses $1, \ldots, n$ according to well established methods implemented in the software package BEAST [32].

**Posterior inference**

Top-level priors for $1/\varphi^2$, $\beta$, $1/\sigma_x^2$, and $1/\sigma_y^2$ are assumed to follow diffuse Gamma$(a, b)$ distributions with $a = 0.001$ and $b = 0.001$. Under the full model, the posterior probability of observing virus and serum locations given immunological data is factored

$$p(X, Y|H) \propto p(H|X, Y, s, v, \varphi) p(X|\beta, \sigma_x, \tau) p(Y|\beta, \sigma_y) p(s, v, \varphi, \beta, \sigma_x, \sigma_y, \tau).$$

We sample from this posterior distribution using the MCMC procedures implemented in the software package BEAST [32]. Metropolis-Hastings proposals include transition kernels that translate individual virus and serum locations $x_i$ and $y_j$ and individual virus effects $v_i$ and serum effects $s_j$, and other transition kernels that scale the entire set of virus and serum locations $X$ and $Y$ and that scale parameters $\varphi$, $\beta$, $\sigma_x$ and $\sigma_y$. For the present analysis, a two-step approach was taken to sample phylogenies, where a posterior sample of phylogenies was gathered using sequence data and then, in the cartographic analysis, trees from this set were randomly proposed and accepted following the Metropolis-Hastings algorithm [33].

**Genetic, antigenic and surveillance data**

We compiled an antigenic dataset of hemagglutination inhibition (HI) measurements of virus isolates against post-infection ferret sera for influenza A/H3N2 by collecting data
from previous publications [6, 8, 9, 11], NIMR vaccine strain selection reports for 2002 and 2008–2012 [34–40] and the Feb 2011 VRBPAC report [41]. We queried the Influenza Research Database [42] and the EpiFlu Database [43] 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 sequence in IRD. Sequences were aligned using MUSCLE v3.7 under default parameters [44]. 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 subsampled 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 402 virus isolates, 519 serum isolates and 10,059 HI measurements. Each virus isolate was compared to an average of 21.9 serum isolates, and each serum isolate was compared to an average of 18.0 virus isolates.

Antigenic data for influenza A/H1N1 was collected from previous publications [8, 11, 45–58] and NIMR vaccine strain selection reports for 2002–2010 [34–37, 59–67]. The same procedure was followed as was followed for H3N2 to match sequence data and to subsample antigenic comparisons. This procedure yielded 115 virus isolates, 77 serum isolates and 1882 HI measurements over the course of 1977 to 2009. Each virus isolate was compared to an average of 10.0 serum isolates, and each serum isolate was compared to an average of 16.2 virus isolates.

Antigenic comparisons for influenza B/Victoria were collated from previous publications [2,8,68–75] and vaccine strain selection reports for 2002–2012 [34–40,59–66,76–78]. Here, the sequence matching and subsampling procedure yielded 179 virus isolates, 70 serum isolates and 2003 HI measurements over the course of 1986 to 2011. Each virus isolate was compared to an average of 6.5 serum isolates, and each serum isolate was compared to an average of 16.7 virus isolates.

Antigenic comparisons for influenza B/Yamagata were collected from previous publications [2,8,68–75,79–85] and vaccine strain selection reports for 2002–2012 [34–40,59–66,76–78]. For B/Yamagata, the matching and subsampling procedure resulted in 174 virus isolates, 69 serum isolates and 1962 HI measurements over the course of 1987 to 2011. Each virus isolate was compared to an average of 6.9 serum isolates, and each serum isolate was compared to an average of 17.3 virus isolates.

Surveillance data was obtained from the Centers of Disease Control and Prevention FluView Influenza Reports from the yearly summaries of influenza seasons 1997–1998 to 2010–2011 [86]. As an example, one report states “collaborating laboratories in the United States tested 195,744 respiratory specimens for influenza viruses, 27,682 (14%) of which were positive. Of these, 18,175 (66%) were positive for influenza A viruses, and 9,507 (34%) were positive for influenza B viruses. Of the 18,175 specimens positive for influenza A viruses, 7,631 (42%) were subtyped; 6,762 (87%) of these were seasonal influenza A (H1N1) viruses, and 869 (13%) were influenza A (H3N2) viruses.” In this case, we estimate the relative proportion of A/H3N2 of the four clades as 0.66 × 0.13 = 0.09. Similar
calculations were performed for A/H1N1, B/Vic and B/Yam.

**Implementation**

Phylogenetic trees were estimated for A/H3N2, A/H1N1, B/Vic and B/Yam using BEAST [32] and incorporated the SRD06 nucleotide substitution model [87], a coalescent demographic model with constant effective population size and a strict molecular clock across branches. MCMC was run for 60 million steps and trees were sampled every 50,000 steps after allowing a burn-in of 10 million steps, yielding a total sample of 2000 trees. These trees were treated as a discrete set of possibilities when subsequently sampled in the BMDS analysis [33]. However, it would be possible to jointly sample from sequence data and serological data using these methods.

MCMC was used to sample virus locations $X$, serum locations $Y$, virus effects $v$, serum effects $s$, MDS precision $1/\varphi^2$, antigenic drift rate $\beta$, virus location precision $1/\sigma^2_x$, serum location precision $1/\sigma^2_y$ and phylogenetic tree $\tau$. MCMC chains were run for 500 million steps and parameter values sampled every 200,000 steps after a burn-in of 100 million steps, yielding a total of 2000 MCMC samples. In all cases when drift parameter $\beta$ was included the MCMC chain mixed well and arrived at the same estimated posterior distribution from different starting points. However, without drift parameter $\beta$, maps for A/H3N2 showed some degree of metastability, where some chains would converge on one solution and other chains would converge on a different solution. We favor models that include $\beta$, because its inclusion, in addition to correcting most identifiability issues, yields much improved mixing of antigenic locations.

There is some difficulty summarizing posterior cartographic samples, as sampled virus and serum locations represent only relative quantities, and because of this, over the course of the MCMC, virus locations may shift. Our prior on virus and serum locations removes much of this issue, orienting the antigenic map along dimension 1 and fixing it to begin at the origin. However, local isometries are often still a problem. For example, in A/H3N2 the HK/68, EN/72 and VI/75 clusters may rotate in relation to other clusters. Consequently, it may be difficult to fully align MCMC samples using Procrustes analysis. For the present study, we take a simple approach and sample a single MCMC step and visualize the antigenic locations at this state (Figures 2, Figure 3). Then, for specific quantities of interest, like rate of antigenic drift and rate of diffusion at different points along the phylogeny, we calculate the quantity across MCMC samples to yield an expectation and a credible interval. This approach accurately characterizes uncertainty that may be hidden in an analysis of a single antigenic map.

**Estimation of diffusion coefficient**

The distance $d$ traveled by a diffusion is not a linear function of time interval $t$, and thus the ‘rate’ of diffusion cannot be calculated following $r = d/t$. Assume we have a one-dimensional diffusion starting from $x_0 = 0$, so that

$$X_t \sim \mathcal{N}(0, \sigma^2 t), \quad (15)$$
where $\sigma$ represents the diffusion volatility, and in this case, the expected squared displacement follows

$$E[X_t^2] = \sigma^2 t. \quad (16)$$

In moving to a two-dimensional diffusion

$$(X_t, Y_t) \sim \mathcal{N}\left(\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix} \sigma_x^2 t & \rho \sigma_x \sigma_y t \\ \rho \sigma_x \sigma_y t & \sigma_y^2 t \end{pmatrix}\right), \quad (17)$$

where $\sigma_x$ represents volatility in dimension 1, $\sigma_y$ represents volatility in dimension 2 and $\rho$ represents correlation, expected squared displacement is equal to the sum of squared displacements in each dimension

$$E[X_t^2 + Y_t^2] = \sigma_x^2 t + \sigma_y^2 t. \quad (18)$$

Such diffusions are often characterized with diffusion coefficient $D$. In this case

$$D = \frac{\sigma_x^2 + \sigma_y^2}{4}, \quad (19)$$

so that expected squared displacement follows from $D$ alone

$$E[X_t^2 + Y_t^2] = 4Dt. \quad (20)$$

Owing to this relationship, Pybus et al. [20] estimate $D$ from phylogenies in which internal nodes have inferred 2D locations via the formula

$$\hat{D} = \frac{1}{n} \sum_{i=1}^{n} \frac{d_i^2}{4t_i}, \quad (21)$$

in which $d_i$ represents the displacement of a branch and $t_i$ represents the length of a branch. However, uncertainty in $d_i$ or $t_i$, especially along short branches of the phylogeny, may cause $d_i^2/4t_i$ to have high variance and increase uncertainty in the resulting estimates of $D$. Consequently, we estimate $D$ by looking at the relationship between total squared displacement and total time along branches of a phylogeny

$$d_\Sigma^2 = \sum_{i=1}^{n} d_i^2$$
$$t_\Sigma = \sum_{i=1}^{n} t_i$$
$$\hat{D} = \frac{d_\Sigma^2}{4t_\Sigma}. \quad (22)$$

This estimator showed less variance than equation 21 when analyzing tree output from the BMDS model and was thus preferred. Additionally, we confirmed through simple simulation that estimator 22 shows lower mean squared error than estimator 21 in the presence of modest levels of noise on estimates of $d_i$ and $t_i$. When estimating $D$ specific to trunk branches or specific to side branches, we calculate $d_\Sigma^2$ and $t_\Sigma$ only from a subset of phylogeny branches as appropriate.
Comparison with previous results

Here, we attempt to compare antigenic locations inferred by our BMDS model to antigenic locations previously inferred by the error minimization methods of Smith et al. [6], referred to here as antigenic cartography by MDS. For this comparison, we use exactly the same HI data used to produce the results in [6], consisting of 273 virus isolates, 79 serum isolates and a total of 4252 HI measurements taken between 1968 and 2003. We begin with a BMDS analog of the antigenic model used in [6], where serum effects are taken as the maximum titer of a particular ferret serum and the expected log₂ drop in HI titer is proportional to Euclidean distance between virus and serum locations. To bring models into further alignment, we use a Uniform(−100, 100) distribution over virus locations and serum locations. Unsurprisingly, we find that this BMDS model produces results that are strongly congruent with MDS results (Figure 8). Antigenic cluster locations are consistent between methods (Figure 8A–B) and antigenic distances between pairs of viruses are consistent between temporally similar and temporally divergent viruses (Figure 8C–E), suggesting that the resulting maps are consistent at both local and global scales. Credible intervals of antigenic distances for the BMDS model remain narrow across the temporal spectrum (Figure 8C–E), implying a fair degree of rigidity to the map.

Smith et al. [6] show that there exist at least two solutions in their assignment of antigenic locations, involving the rotation of clusters HK/68, EN/72 and VI/75 (shown in Figure S2 of [6]). We observe the same metastable behavior in our analysis; some MCMC chains converge on the solution shown in Figure 8B, while other MCMC chains converge on the alternative solution shown in Figure 9B. The distribution of likelihood values appear highly similar between these two solutions, suggesting that they represent global optima. The rotation of the HK/68, EN/72 and VI/75 clusters creates a map that bends slightly, so that temporally distant viruses appear closer in the rotated solution than in the original solution (Figure 9C–E). In this case, it is clear that the solutions are locally consistent between viruses up to ~15 years divergent, even if there some degree of global flexibility.

As discussed in the main text, the presence of multiple optima with different degrees of 2D curvature implies an issue of identifiability; the HI likelihood model alone cannot distinguish between these possibilities. Because of this issue, and to more easily estimate rates of antigenic drift, we include a model of systematic drift in antigenic location that favors linear movement in the antigenic map. We find that including this drift prior on antigenic locations removes the problem of identifiability. Antigenic locations produced by this model remain locally consistent with MDS results between viruses ~15 years divergent, but global comparisons show that this BMDS model has partitioned more variance to the first antigenic dimension (Figure 10). We additionally find that including the drift prior on antigenic locations often results in greater predictive power, with a slight improvement of test error for the A/H1N1, B/Vic and B/Yam datasets (Table 1).

Our final BMDS model (model 9, Table 1) differs from antigenic model used by Smith et al. [6] in including temporally- and phylogenetically-informed priors on antigenic locations and also in estimating serum and virus effects. Here, we investigate the impact on antigenic locations of estimating virus and serum effects in the BMDS model. To isolate this difference, we use a Uniform(−100, 100) prior on antigenic locations. Surprisingly, estimating
Figure 8. Comparison of A/H3N2 antigenic locations estimated by Smith et al. [6] using MDS and an equivalent BMDS model. (A) MDS antigenic locations, reoriented so that the primary dimension lies on the x-axis rather than on the y-axis as in Figure 1 of [6]. (B) A posterior sample of antigenic locations from an equivalent BMDS model. In (A) and (B), viruses are shown as colored circles, with color denoting antigenic cluster inferred by [6], and sera are shown as gray points. (C) Antigenic distances between A/Bilthoven/16190/1968 and all other viruses determined for both methods. (D) Antigenic distances between A/Fujian/411/2002 and all other viruses determined for both methods. (E) Antigenic distances between 750 random pairs of viruses determined for both methods. In (C), (D) and (E) red points show distances for the MDS model and gray bars show the 95% credible interval of distances for the BMDS model, while the red dashed line shows a LOESS regression to MDS distances, and the black dashed line shows a LOESS regression to the BMDS distances. The BMDS model has a Uniform(−100, 100) prior on antigenic locations and serum effects fixed at maximum titer values.

virus and serum effects results in a more linear antigenic map (Figure 11), resembling the appearance of the map incorporating the antigenic drift prior, while preserving local consistency. We generally observe congruence between MDS and BMDS antigenic locations for viruses less than ~10 years divergent (Figure 11E). However, specific viruses may be affected, for instance A/Bilthoven/16190/1968 (Figure 11C), which appears more distant from all other viruses when serum and virus effects are included.

In this dataset, viruses 15 or more years divergent always yield threshold titers, and hence, their relative locations must be indirectly inferred rather than through direct comparison. This may explain why we observe local consistency between models at scales less than ~15 years, but some degree of global inconsistency. Still, these results suggest that, when making local comparisons, such as those used to calculate year-to-year antigenic drift (Figure 3), outcomes are expected to be robust to many model particulars.
Figure 9. Comparison of A/H3N2 antigenic locations estimated by Smith et al. [6] using MDS and an equivalent BMDS model under an alternative solution. (A) MDS antigenic locations, reoriented so that the primary dimension lies on the x-axis rather than on the y-axis as in Figure 1 of [6]. (B) A posterior sample of antigenic locations from an equivalent BMDS model that has converged on the alternative solution. In (A) and (B), viruses are shown as colored circles, with color denoting antigenic cluster inferred by [6], and sera are shown as gray points. (C) Antigenic distances between A/Bilthoven/16190/1968 and all other viruses determined for both methods. (D) Antigenic distances between A/Fujian/411/2002 and all other viruses determined for both methods. (E) Antigenic distances between 750 random pairs of viruses determined for both methods. In (C), (D) and (E) red points show distances for the MDS model and gray bars show the 95% credible interval of distances for the BMDS model, while the red dashed line shows a LOESS regression to MDS distances, and the black dashed line shows a LOESS regression to the BMDS distances. The BMDS model has a Uniform(−100, 100) prior on antigenic locations and serum effects fixed at maximum titer values.

Availability

Source code implementing the cartographic models has been made fully available as part of the software package BEAST [32], and can be downloaded from its Google code repository (http://code.google.com/p/beast-mcmc/).

Acknowledgments

We thank Richard Reeve, Dan Haydon, and Simon Frost for insights on antigenic modeling and MDS. We acknowledge the laboratories that provided sequences to EpiFlu database: Centers for Disease Control and Prevention (USA), Chinese Center of Disease Prevention and Control, Hospital Clinic of Barcelona, National Institute of Hygiene of Morocco, National Institute of Infectious Diseases (Japan), National Institute for Medical Research (UK), Norwegian Institute of Public Health, Swedish Institute for Infectious Disease Control, Victorian Infectious Diseases Reference Laboratory (Australia).
Figure 10. Comparison of A/H3N2 antigenic locations estimated by Smith et al. [6] using MDS and an extended BMDS model that includes date-informed priors on antigenic locations. (A) MDS antigenic locations, reorientated so that the primary dimension lies on the $x$-axis rather than on the $y$-axis as in Figure 1 of [6]. (B) A posterior sample of antigenic locations from a BMDS model that includes date-informed priors on antigenic locations. In (A) and (B), viruses are shown as colored circles, with color denoting antigenic cluster inferred by [6], and sera are shown as gray points. (C) Antigenic distances between A/Bilthoven/16190/1968 and all other viruses determined for both methods. (D) Antigenic distances between A/Fujian/411/2002 and all other viruses determined for both methods. (E) Antigenic distances between 750 random pairs of viruses determined for both methods. In (C), (D) and (E) red points show distances for the MDS model and gray bars show the 95% credible interval of distances for the BMDS model, while the red dashed line shows a LOESS regression to MDS distances, and the black dashed line shows a LOESS regression to the BMDS distances. The BMDS model has a date-informed prior on antigenic locations and serum effects fixed at maximum titer values.

Author contributions

AR, MAS, TB conceived the study. AR, MAS, TB, DJS, PL designed the antigenic model. TB, GD, VG, JWM, AJH, CAR, DJS, AR gathered antigenic and genetic data. AR, MAS, TB, PL implemented statistical phylogenetic procedures. TB performed the analysis. TB, AR, DJS, JWM, PL, CAR, AJH, GD, MAS interpreted the results. TB, MAS, AR, JWM, AJH wrote the paper.

Funding

TB was supported by a Newton International Fellowship from the Royal Society. MAS was supported by National Institutes of Health grants R01 GM086887 and R01 HG006139 and National Science Foundation grant DMS0856099. The research leading to these results has received funding from the European Research Council under the European Community’s Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no. 278433-PREDEMICS and ERC Grant agreement no. 260864. VG, AJH and JWM were funded
Figure 11. Comparison of A/H3N2 antigenic locations estimated by Smith et al. [6] using MDS and an extended BMDS model that estimates serum and virus effects. (A) MDS antigenic locations, reorientated so that the primary dimension lies on the x-axis rather than on the y-axis as in Figure 1 of [6]. (B) A posterior sample of antigenic locations from a BMDS model that estimates virus and serum effects. In (A) and (B), viruses are shown as colored circles, with color denoting antigenic cluster inferred by [6], and sera are shown as gray points. (C) Antigenic distances between A/Bilthoven/16190/1968 and all other viruses determined for both methods. (D) Antigenic distances between A/Fujian/411/2002 and all other viruses determined for both methods. (E) Antigenic distances between 750 random pairs of viruses determined for both methods. In (C), (D) and (E) red points show distances for the MDS model and gray bars show the 95% credible interval of distances for the BMDS model, while the red dashed line shows a LOESS regression to MDS distances, and the black dashed line shows a LOESS regression to the BMDS distances. The BMDS model has a Uniform(−100, 100) prior on antigenic locations and virus and serum effects estimated in a hierarchical Bayesian fashion.

by MRC reference no U117512723. AR acknowledges the support of the Wellcome Trust (grant no. 092807). Collaboration between MAS, AR and PL was supported by the National Evolutionary Synthesis Center (NESCent), NSF EF-0423641. DJS and CAR acknowledge EU FP7 programs EMPerIE (223498) and ANTIGONE (278976), Human Frontier Science Program (HFSP) program grant P0050/2008, Wellcome 087982/A1A, NIH Director’s Pioneer Award DP1-OD000490-01, and National Institute of Allergy and Infectious Diseases NIH CEIRS contract HHSN266200700010C. CAR was supported by a University Research Fellowship from the Royal Society.

References


26
sites of Hong Kong influenza haemagglutinin and their involvement in antigenic

8: 196–205.

5. Hirst G (1943) Studies of antigenic differences among strains of influenza A by

376.


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

mendations for the viruses to be used in the 2012 Southern Hemisphere Influenza
Vaccine: Epidemiology, antigenic and genetic characteristics of influenza A (H1N1)
pdm09, A (H3N2) and B influenza viruses collected from February to September

13. Goldstein E, Cobey S, Takahashi S, Miller J, Lipsitch M (2011) Predicting the
epidemic sizes of influenza A/H1N1, A/H3N2, and B: a statistical method. PLoS
Med 8: e1001051.

14. Ferguson NM, Galvani AP, Bush RM (2003) Ecological and immunological deter-

genic drift in the influenza A virus (H3N2) nucleoprotein and escape from recogni-

in seasonal activity of influenza A/H3N2, A/H1N1, and B from 1997 to 2005: viral


41. 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.


