Identification of Low- and High-Impact Hemagglutinin Amino Acid Substitutions That Drive Antigenic Drift of Influenza A (H1N1) Viruses

William T. Harvey[1], Donald J. Benton[2], Victoria Gregory[2], James P. J. Hall[3¤], Rodney S. Daniels[2], Trevor Bedford[4], Daniel T. Haydon[1], Alan J. Hay[2], John W. McCauley[2], Richard Reeve[1,5]*

1 Boyd Orr Centre for Population and Ecosystem Health and Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom, 2 The Crick Worldwide Influenza Centre, The Francis Crick Institute, Mill Hill Laboratory, The Ridgeway, Mill Hill, London, United Kingdom (formerly WHO Collaborating Centre for Reference and Research on Influenza, Division of Virology, MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London, United Kingdom), 3 Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom, 4 Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America, 5 The Pirbright Institute, Pirbright, Woking, Surrey, United Kingdom

¤ Current address: Department of Biology, University of York, York, United Kingdom
* richard.reeve@glasgow.ac.uk

Abstract

Determining phenotype from genetic data is a fundamental challenge. Identification of emerging antigenic variants among circulating influenza viruses is critical to the vaccine virus selection process, with vaccine effectiveness maximized when constituents are antigenically similar to circulating viruses. Hemagglutination inhibition (HI) assay data are commonly used to assess influenza antigenicity. Here, sequence and 3-D structural information of hemagglutinin (HA) glycoproteins were analyzed together with corresponding HI assay data for former seasonal influenza A(H1N1) virus isolates (1997–2009) and reference viruses. The models developed identify and quantify the impact of eighteen amino acid substitutions on the antigenicity of HA, two of which were responsible for major transitions in antigenic phenotype. We used reverse genetics to demonstrate the causal effect on antigenicity for a subset of these substitutions. Information on the impact of substitutions allowed us to predict antigenic phenotypes of emerging viruses directly from HA gene sequence data and accuracy was doubled by including all substitutions causing antigenic changes over a model incorporating only the substitutions with the largest impact. The ability to quantify the phenotypic impact of specific amino acid substitutions should help refine emerging techniques that predict the evolution of virus populations from one year to the next, leading to stronger theoretical foundations for selection of candidate vaccine viruses. These techniques have great potential to be extended to other antigenically variable pathogens.

Author Summary

Influenza A viruses are characterized by rapid antigenic drift: structural changes in B-cell
epitopes that facilitate escape from pre-existing immunity. Consequently, seasonal influenza continues to impose a major burden on human health. Accurate quantification of the
antigenic impact of specific amino acid substitutions is a pre-requisite for predicting the
fitness and evolutionary outcome of variant viruses. Using assays to attribute antigenic
variation to amino acid sequence changes we identify substitutions that contribute to antigenic drift and quantify their impact. We show that substitutions identified as low-impact
are a critical component of virus antigenic evolution and by including these, as well as the
high-impact substitutions often focused on, the accuracy of predicting antigenic phenotypes of emerging viruses from genotype is doubled.

Introduction

Antigenic evolution of human influenza A viruses is characterized by rapid drift, with structural changes in antigenic epitopes allowing the virus to escape existing immunity. Consequently, seasonal influenza continues to impose a major burden on human health causing
250,000 to 500,000 deaths annually [1]. Influenza vaccines, which remain the most effective
means of disease prevention, currently comprise antigens from A(H1N1), A(H3N2) and B
viruses predicted to elicit the most effective immune responses against circulating viruses in
the forthcoming influenza season [1,2]. The continually evolving antigenic phenotype of influenza A viruses presents an ongoing challenge for vaccine virus selection, as effectiveness is
greatest when vaccine components are antigenically similar to circulating viruses. The HA glycoprotein is the key antigenic determinant of influenza viruses [3] and consequently the most
critical vaccine component. Neutralizing antibodies primarily bind to amino acids in protruding loops and helices that form defined antigenic sites [4,5], and substitutions that alter epitope
structure can inhibit antibody binding and help the virus escape existing immunity [6]. When
the selective advantage conferred is sufficient, novel antigenic variants will replace circulating
viruses. Hence, phylogenies of influenza A HA sequences are characterized by the presence of a
single predominant trunk lineage, and short side branches, representing the rapid turnover of
the influenza virus population [7,8].

Antigenic changes in circulating influenza viruses are principally assessed by the HI assay
[9,10]. Results of many HI assays can be summarized using cartographic approaches, which
approximate antigenic dissimilarity by Euclidean distances between viruses and antisera on a
map, with antigenic evolution in influenza represented as movement between clusters of
viruses [11]. The non-synonymous genetic mutation(s) causing transitions between antigenic
clusters can be determined experimentally by reverse genetics [12], though this approach is
often laborious, as multiple amino acid substitutions bridge each antigenic cluster transition,
and individual substitutions need to be assessed. This approach recently demonstrated that
transitions between antigenic clusters of H3N2 viruses are caused predominantly by single
amino acid substitutions at positions near the receptor-binding site [12]. However, major cluster transitions may not be the only antigenically important events [13,14] and an exhaustive
reverse genetics analysis of all observed substitutions is not feasible due to high levels of amino
acid sequence diversity in HA (e.g. at 46% of amino acid positions, in this study).

An alternative approach is to integrate matching sequence, antigenic and 3-D structural
data into models that allow us to attribute the observed antigenic differences in a dataset
directly to their underlying causes. Reeve *et al.* developed such a model to identify surface-

exposed regions of the capsid proteins of foot-and-mouth disease virus where substitutions were correlated with antigenic change, but were unable to show definitive causal connection with specific substitutions [15]. Various other computational approaches have similarly been used to identify antigenically important amino acid positions in influenza HA by comparison of predominant sequences of successive antigenic clusters and by comparing sequence and antigenic data [11,16–19].

In this paper, we: 1. extend the modeling approach of Reeve *et al.* [15] to former seasonal influenza A(H1N1) viruses (*i.e.* A(H1N1) viruses circulating in humans prior to the 2009 pandemic), focusing on these rather than A(H3N2) viruses, for which the role of neuraminidase-mediated agglutination of red blood cells (RBCs) has complicated the relationship between HI data and antigenic change [20], or the distinct A(H1N1)pdm09 viruses, which have largely remained antigenically similar since emerging in humans in 2009 [21]; 2. attribute variation in HI titers to individual amino acid substitutions; 3. quantify their antigenic impact; 4. assess, by reverse genetics, the impact of a subset of the identified substitutions to validate the model; 5. show how inferences, based on the determinants of low-impact and high-impact antigenic changes, improve our understanding of the antigenic evolution of the virus; and 6. demonstrate that the characterization of these antigenic determinants allows us to accurately assess directly from HA gene sequence data the antigenicity of newly emerging viruses, measurement of which is critical to predicting the evolutionary success of newly emerging variants.

Results

We compiled HI assay data from former seasonal A(H1N1) viruses isolated between 1997 and 2009, comprising 19,905 individual measurements of cross-reactivity between 43 post-infection ferret antisera and 506 viruses. The HA1 sequences for all 506 viruses tested by HI in this dataset were used to generate a maximum clade credibility tree [22]. Each of these 506 viruses was present in the HI dataset as test viruses, being tested against various antisera using the HI assay. Of these 506 viruses, 43 were also selected as reference viruses and used to raise antisera used in this HI dataset. HA1 trees for the complete set of 506 viruses and for the 43 reference strains only are shown with corresponding HI titers represented as a heat-map in Fig 1.

In Fig 1 viruses and antisera are sorted phylogenetically according to the maximum clade credibility trees and colored cells indicate average HI titres for pairs of virus and antiserum tested. It is clear from this figure that, generally, viruses that are phylogenetically related are also antigenically similar, however there are instances where phylogenetically similar viruses are antigenically distinct. For example, the starkest antigenic change is represented as the red to yellow change in the columns for the antisera raised against reference viruses A/Johannesburg/82/96 and A/Bayern/7/95 to the left of the heat-map in Fig 1. The color change in these columns of the heat-map corresponds to a relatively deep bifurcation in the HA1 phylogeny to the left of Fig 1, where one or more changes in amino acid must have caused a disproportionally large change in antigenic structure. This demonstrates the level of heterogeneity in the antigenic impact of genetic changes that exists.

The effect of amino acids at specific positions

To identify antigenic relationships and their predictors, we used linear mixed effects models to account for variation in the HI titers, as described by Reeve *et al.* [15]. Initial model selection identified non-antigenic sources of variation in HI titer. We determined that a fixed effect, a_v, for each of the 506 viruses tested, v, should be included in the model (Likelihood ratio test (LRT), $p < 10^{-20}$), to account for consistent differences in titers between viruses, reflecting changes in receptor-binding avidity amongst other factors. A further fixed effect, s_r, was
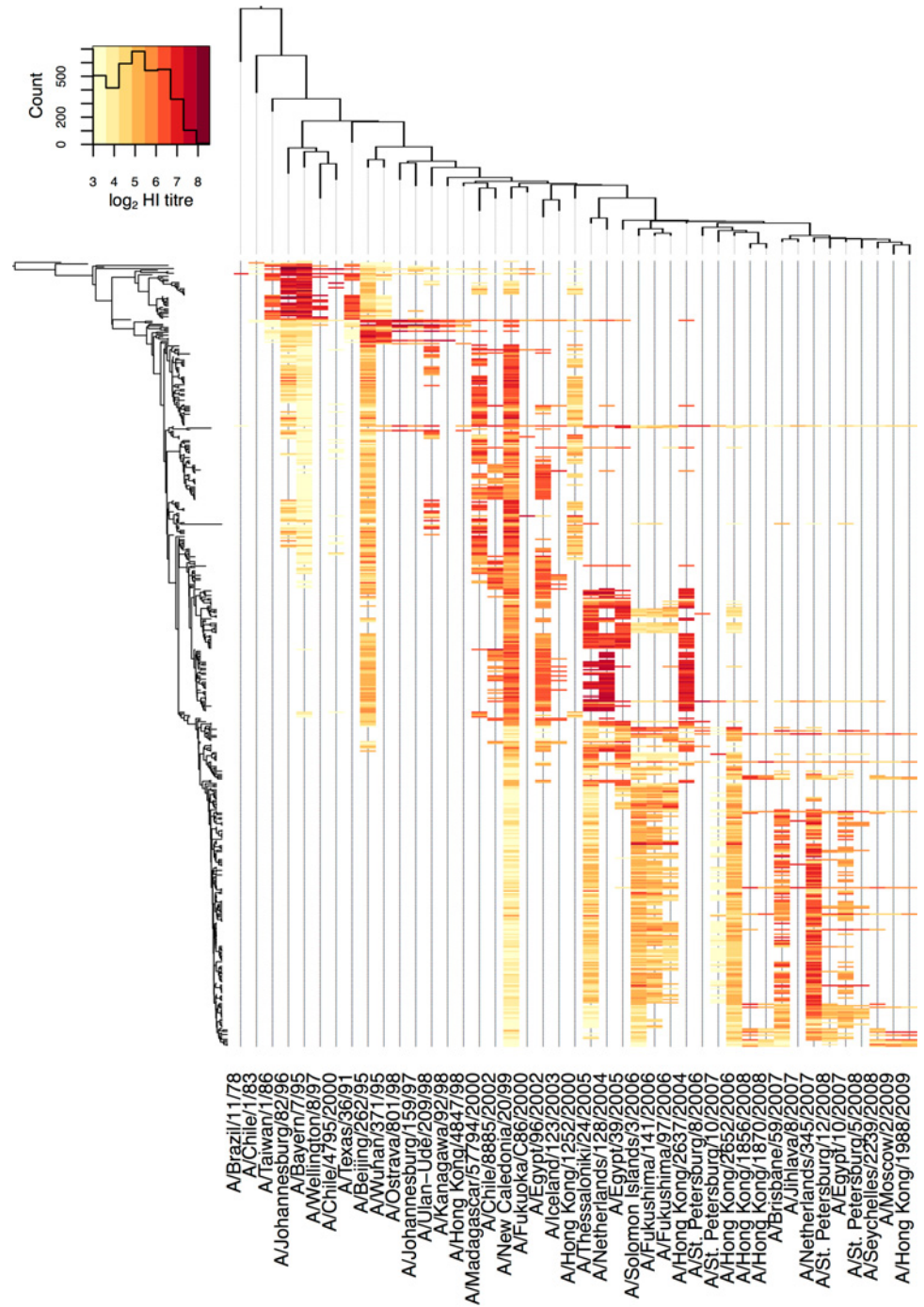
Fig 1. Heat-map illustrating the relationship between molecular and antigenic evolution. Cells are colored by mean log HI titer for each pairing of antiserum and test virus present in the full dataset. Test viruses and reference viruses used to generate post-infection ferret antisera are sorted phylogenetically on the HA gene along the vertical and horizontal axes respectively. Phylogenies are shown to the left for test viruses and above for reference viruses. The color key for HI titers is shown in the histogram at top left along with the number of assays yielding each titer. S1 Fig provides examples of the observed variability in HI titer for the most frequently used virus.

doi:10.1371/journal.ppat.1005526.g001

required for each of the 43 reference viruses, r (LRT, p<10^{-20}), to account for consistent differences in titers between antisera raised against different reference viruses, potentially reflecting differences in immunogenicity. Date of test needed to be controlled for as a random effect, ε_D, with groups for the 351 dates on which data used were collected accounting for variability in batches of RBCs and dilutions of RBCs, antisera and viruses. Improvements in AIC are shown in S1 Table. These factors compensate for non-antigenic effects impacting HI titers (Eq 1).

$$\log_2(H_{r,v}) = k_0 + k_j\alpha_j(r,v) + s_r + a_v + \varepsilon_D + \varepsilon_R \tag{1}$$

$H_{r,v}$ is the HI titer for test virus v and antiserum raised against reference virus r. k_0 is a baseline, and ε_R is the residual measurement error not explained by the model.

Eq 1 includes a term, $k_j\alpha_j(r,v)$, to investigate the effect of amino acid substitutions at specific positions: α_j represents the presence (1) or absence (0) of substitution at a specific amino acid position between the reference virus, r, and test virus, v, and k_j is the associated regression coefficient. Using this model (Eq 1) substitutions at over 50% of non-conserved, surface-exposed positions and over 25% of non-conserved, non-surface-exposed positions were significantly correlated with reduced HI titer (LRT, p<0.05) using a Holm-Bonferroni correction for multiple tests [23]. Furthermore, the number of synonymous mutations between viruses was significantly correlated with reduced titer (LRT, p<10^{-15}) because of a correlation between molecular and antigenic evolution that arises due to the hitchhiking of neutral mutations on beneficial backgrounds. This demonstrates that a simple regression analysis will incorrectly identify some antigenically neutral changes as antigenically important–*i.e.* false positives–simply because they occur at a similar point in the evolutionary history of the virus to one or more antigenically important substitutions (*i.e.* in the same, or a nearby branch of the phylogeny).

Incorporating phylogenetic structure

The described tendency for identification of false positives required phylogenetic structure to be reflected in the model. Eq 7 of Reeve *et al.* [15] was used to identify branches of the phylogeny that were correlated with lower HI titers when they separated reference virus and test virus:

$$\log_2(H_{r,v}) = k_0 + \sum_i m_i\delta_i(r,v) + s_r + a_v + \varepsilon_D + \varepsilon_R \tag{2}$$

Eq 2 incorporates branch terms $m_i\delta_i(r,v)$ instead of the term representing substitutions at a single amino acid position: $\delta_i = 1$ when reference virus (r) and test virus (v) are separated by branch i of the phylogeny and $\delta_i = 0$ otherwise, with m_i being the associated regression coefficient from the mixed effects model. The tree generated for the 506 viruses in our dataset contained 1010 branches and it was computationally unfeasible to search the 2^{1010} possible antigenically important sets of branches so a stochastic hill-climbing approach was used to identify 62 branches as correlating with drops in HI titer when they separated reference and test viruses, indicating that antigenic evolution occurred in these branches. Such antigenic events were found in much higher proportion in the trunk (38.3%) than in side (4.6%) branches (χ^2 test, p<10^{-13}), supporting the standard model of influenza antigenic drift, whereby substitutions altering antigenicity without loss of fitness undergo preferential fixation, thus forming the trunk lineage from which future viruses descend [7,8]. Including these phylogenetic branch terms improved model fit substantially, reducing AIC by 26,969 (S1 Table). With these 62 branch terms included in the model (Eq 2), there was no longer a significant correlation between HI titer and the number of synonymous mutations between reference and test virus (LRT, p>0.2), nor with any of the non-surface-exposed positions (LRT, p>0.05). This

shows that including branch terms accounted for the antigenic variability in the data and reduced the false discovery rate as expected.

Substitutions affecting antigenicity in multiple positions of the phylogeny

The model was extended by combining Eqs 1 and 2 to include explicit terms for amino acid substitutions and branch terms, as in Eq 8 of Reeve *et al.* [15], to identify antigenically important substitutions:

$$\log_2(H_{r,v}) = k_0 + k_j\alpha_j(r,v) + \sum_{i=1}^{62} m_i\delta_i(r,v) + s_r + a_v + \varepsilon_D + \varepsilon_R \qquad (3)$$

Eq 3 incorporates the term $k_j\alpha_j(r,v)$ from Eq 1 representing substitution at specific amino acid positions. Each of the previously identified 62 branch terms (δ_i) were included and associated regression coefficients (m_i) were re-estimated in a model containing the $k_j\alpha_j(r,v)$ term. Because branch terms account for the antigenic changes inferred to occur in single specific branches of the phylogeny, any significant improvement to model fit by α_1 is a result of the term representing amino acid substitution at a particular HA1 position being correlated with a change in the antigenicity of the virus represented in multiple branches of the phylogeny. Thus an improvement to model fit achieved by inclusion of α_1 indicates that there have been alternative, convergent- or back-substitutions at the same amino acid position associated with antigenic change in at least two branches of the phylogeny.

We investigated 110 non-conserved, surface exposed amino acid positions of the HA1 domain. At four of these (141, 153, 187 and 190), the inclusion of an α_1 term representing substitution (Eq 3) improved model-fit compared with the model containing only branch terms (Eq 2). Since the identified positions improve model fit in the presence of branch terms (δ_i), we can infer that substitutions at these positions correlate with antigenic change in more than one position of the phylogeny. We describe substitutions at these positions as having *support across the phylogeny*. Each of these four amino acid positions (Fig 2A) has previously been allocated to one of the H1 antigenic sites [5]. Position 187 is also a constituent of the primary sialic acid receptor-binding site and the analogous position 190 in H3-HA has been described as forming hydrogen bonds with the 9-hydroxyl group of sialic acid [3]. Non-surface exposed positions were examined separately; substitution at none of these positions improved model fit indicating that the classification of surface-exposed and non-surface-exposed positions did not have implications for these analyses, as they would not have been selected as antigenically important.

Different substitutions at the same position are expected to vary in antigenic impact according to the biochemical properties of the amino acids involved. To account for this, we measured the significance and average impact (in antigenic units, where a unit corresponds to a 2-fold dilution in the HI assay) of each substitution at HA1 positions 141, 153, 187 and 190 that was observed to have occurred between reference and test viruses in the dataset. Substitutions between seven pairs of amino acids at the four positions showed significant antigenic impact with *support across the phylogeny*. The mean antigenic impact (k_j in Eq 3) of exchange between amino acids of each pair is shown in Table 1.

Substitutions affecting antigenicity at single positions in the phylogeny

Next, we added terms k_j and α_j to represent each of the seven inferred antigenic substitutions at the four positions with support across the phylogeny (Table 1) to produce Eq 4. We then investigated the causes of antigenic change in branches that still had large estimated antigenic
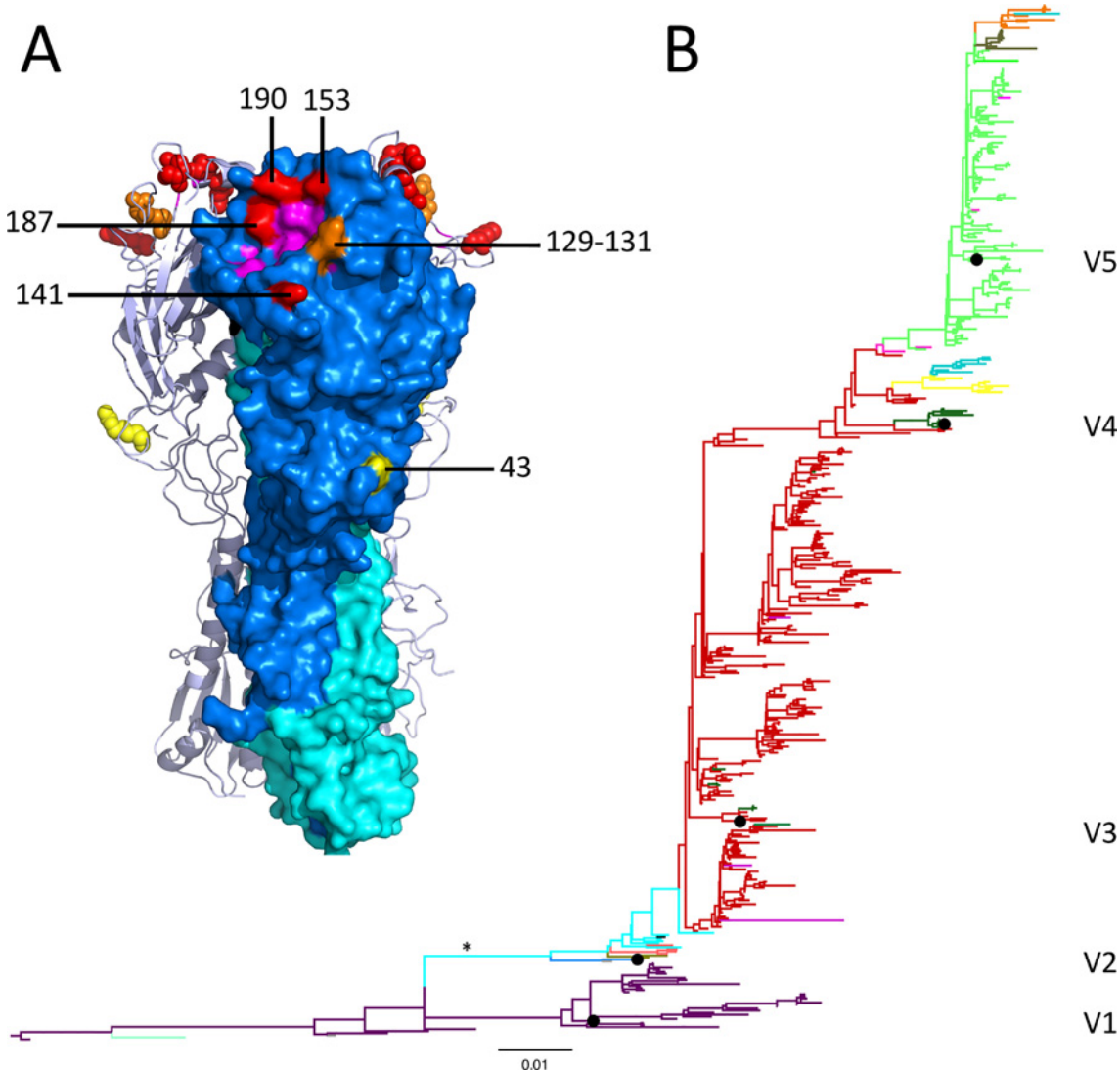
Fig 2. HA positions implicated in antigenic evolution and locations of associated substitutions in HA1 phylogeny. (A) Amino acid positions that affect antigenic phenotype modeled on the HA structure of A/Puerto Rico/8/34 (Protein Data Bank ID: 1RU7) [24]. Surface representation of the front monomer is shown with HA2 in cyan and HA1 in blue with residues of the receptor-binding site colored pink. Positions with substitutions that can explain antigenic change in multiple locations across the phylogeny are shown in red. Residues adjacent to the position of the K130 deletion are colored orange with the locations of the co-occurring R43L, F71I and S271P substitutions are colored yellow. Residues are labeled on the front HA1 monomer and shown as spheres on the remaining backbones. (B) HA1 phylogeny showing positions of significant antigenic substitutions. Color changes mark the locations of substitutions associated with changes in antigenic phenotype of at least 0.5 antigenic units. The position of the branch associated with the greatest drop in cross-reactivity is marked (*). Black circles indicate the positions of viruses included in the influenza vaccine over the period of HI data collection and are labeled: A/Bayern/7/95 (V1), A/Beijing/262/95 (V2), A/New Caledonia/20/99 (V3), A/Solomon Islands/3/2006 (V4) and A/Brisbane/59/2007 (V5). Branch length indicates the estimated number of nucleotide substitutions per site.

impacts.

$$\log_2(H_{r,v}) = k_0 + k'_i \alpha'_l(r,v) + \sum_{j=1}^{7} k_j \alpha_j(r,v) + \sum_{i=1}^{61} m_i \delta_i(r,v) + s_r + a_v + \varepsilon_D + \varepsilon_R \quad (4)$$

Terms for these seven substitutions absorb variation in HI previously explained by branch terms that correspond to the positions in the phylogeny where those substitutions were estimated to have occurred. However, this model still included 18 branch terms representing internal branches of the phylogeny whose estimated impact, m_i, on the HI assay remained

Table 1. HA1 amino acid substitutions that correlate with antigenic change.

Substitution(s) (H1-HA numbering)	Antigenic site		Antigenic impact * (antigenic units)
	H1[5]	H3[4]	
Substitutions with support across phylogeny identified using Eq 3†:			
K141E	Ca	A	2.37 (2.27–2.47)
E153G	Sa	B	0.20 (0.07–0.33)
E153K	Sa	B	0.66 (0.39–0.93)
G153K	Sa	B	1.50 (0.51–2.49)
D187N	Sb	B	0.33 (0.30–0.36)
D187V	Sb	B	0.88 (0.51–2.49)
A190T	Sb	B	0.24 (0.17–0.31)
Substitutions without support across phylogeny identified using Eq 4†:			
S36N		C	0.66 (0.22–1.11)
S72F	Cb	E	0.81 (0.49–1.13)
E74G, E120G‡	Cb,-	E,A	0.43 (0.29–0.57)
R43L, F71I, ΔK130, S271P‡	-,Cb,-,-	C,-,A,-	3.53 (3.44–3.62)
S142N	Ca	A	0.75 (0.58–0.92)
K163N	Sa		0.67 (0.62–0.73)
S183P		B	0.61 (0.33–0.89)
N184S	Sb	B	0.51 (0.31–0.70)
W252R			0.37 (0.32–0.43)
E274K			1.31 (0.68–1.93)
R313K			1.47 (0.84–2.10)

* k_j in Eq 3 or k_i' in Eq 4. Mean and 95% CI are shown.

† Substitutions identified by likelihood ratio test using p-value of 0.05 adjusted using Bonferroni correction.

‡ Multiple substitutions in the same branch offer alternative explanations for the associated antigenic change.

detectable (at least 0.25 antigenic units) in the model containing terms for each of the seven substitutions. Each of these 18 branch terms were excluded in turn, the model re-built with the 61 residual branches, and each remaining amino acid position (as k' and α') was retested to determine which substitution(s) could explain the variation in HI titer associated with the excluded branch term. A substitution identified at this stage (when a branch term had been excluded) was inferred to have caused the associated antigenic change at that position in the phylogeny if it was the only substitution identified.

In nine cases, a single substitution was identified as explaining variation in HI titer upon exclusion of one branch. These substitutions were at positions 36, 72, 142, 163, 183, 184, 252, 274 and 313 (Table 1). Unique identification was not possible in two further cases, as multiple substitutions occurring in the same branch could not be discriminated. The branch associated with the greatest drop in HI titer (-3.61 units) across the phylogeny (starred in Fig 2B) has a deletion of lysine at position 130 (ΔK130) and substitutions R43L, F71I and S271P. The antigenic significance of ΔK130 has been described [25]; however each of these co-occurring substitutions have been identified as antigenic determinants by another *in silico* technique, which did not identify them as false positives [17]. Each of these four changes is assigned equal weight in our model but we identify explicitly that they offer alternative explanations for the same antigenic change and are not independent antigenic determinants. To infer their individual effects experimental investigation was required. One further instance of alternative

substitutions at different positions explaining an antigenic change equally well involved positions 74 and 120. The nine single substitutions and these two instances where alternative substitutions explained antigenic change equally well gave a total of eleven cases where antigenic change in a single position of the phylogeny could be attributed to an amino acid substitution, or multiple substitutions (Table 1).

Although the substitutions identified when branch terms at these eleven positions of the phylogeny were excluded correlated with antigenic change at only a single position in the phylogeny, it is notable that, among them, positions 72, 74, 142, 163 and 184 map to previously described H1 antigenic sites while positions analogous to 36, 120 and 183 are constituents of H3 antigenic sites [4]. Locations within the phylogeny where any of the identified substitutions in Table 1 altered the antigenic phenotype of the virus by at least 0.5 antigenic units and the degree of correspondence with changes to the H1 vaccine component are shown in Fig 2B. Each of the five vaccine components in this phylogeny are separated by at least one color change indicating that potential genetic drivers for all of the most important antigenic changes in the period studied have been identified.

Production of mutant viruses by reverse genetics

To validate the identification of substitutions affecting antigenicity and assess the accuracy of estimated antigenic effects mutant viruses containing a subset of the amino acid substitutions identified in Table 1 were generated by reverse genetics. The HA gene of an exclusively cell culture-propagated virus, A/Netherlands/1/93 (Neth93), was used. We introduced the K130 deletion (ΔK130) and the R43L substitution into the Neth93 HA independently to test whether both of these changes cause antigenic change. Given the large antigenic impact of ΔK130 [25], its introduction generated an additional, antigenically distinct HA background (Neth93 Δ130) in which to further test the effects of other substitutions (Table 1): the HA genes of both Neth93 and Neth93 Δ130 were used to produce viruses carrying individual substitutions of K141E, E153K and D187N.

Mutant recombinant viruses were characterized by HI using a panel of post-infection ferret antisera raised against seven reference viruses with HA1 amino acid identity at positions 43, 130, 141, 153 and 187 as shown in S3 Table. To assess the antigenic impact of each amino acid substitution introduced by mutagenesis, antisera of two types were chosen: 1) antisera raised against parent-like viruses that had amino acid identity in common with the parent virus (*i.e.* R43 for the substitution R43L); 2) antisera raised against mutant-like viruses that had amino acid identity in common with the recombinant virus (*i.e.* L43 for the substitution R43L). The mean change in $\log_2$ HI titre (H), associated with the introduction of each substitution, was averaged across antisera assigned to each of these two groups independently (ΔH_1 and ΔH_2 respectively). For each amino acid substitution introduced into the recombinant viruses, the assignment of antisera to these two categories, based on reference virus amino acid sequence, is shown in red or blue cell color respectively in S3 Table.

Estimating the antigenic and non-antigenic effects of introduced substitutions

In addition to antigenic change, HI titers can be affected by variation in other properties of the test virus, notably receptor-binding avidity, and individual amino acid substitutions may cause fluctuation in HI titer as a result of variation in these other properties [26,27]. By using antisera raised against parent-like and mutant-like viruses, changes in $\log_2$ HI titer between parent and mutant recombinant viruses resulting from antigenic (ΔH_A) and non-antigenic (ΔH_N) effects

Table 2. Comparison of predicted and observed antigenic impacts of HA1 amino acid substitutions assessed by HI.

Substitution	Predicted	Mutagenesis	Observed effect[†]	
	antigenic effect*	Background	ΔH_A	ΔH_N
K141E[‡]	2.37	Neth93 Δ130	2.60	+0.27
E153K	0.66	Neth93	0.67	-0.42
		Neth93 Δ130	0.65	-2.15
		Averaged	0.66	-1.28
D187N	0.33	Neth93	0.41	-0.41
		Neth93 Δ130	-0.08	-0.54
		Averaged	0.16	-0.47
ΔK130	3.53[§]	Neth93	4.10	-0.78
R43L	3.53[§]	Neth93	0.01	-0.01

* Predicted mean antigenic impact (from Table 1) measured in antigenic units.

[†] Mean observed changes in $\log_2$ HI titer (in antigenic units) partitioned into antigenic (ΔH_A) and non-antigenic (ΔH_N) effects.

[‡] HA plasmid was generated for the mutant Neth93 K141E but multiple attempts to rescue virus were unsuccessful.

[§] ΔK130 and R43L occur in the same branch of the phylogeny offering alternative explanations for the associated antigenic change.

doi:10.1371/journal.ppat.1005526.t002

could be distinguished using Eq 5.

$$\Delta H_A = \frac{(\Delta H_2 - \Delta H_1)}{2} \qquad \Delta H_N = \frac{(\Delta H_1 + \Delta H_2)}{2} \tag{5}$$

If the amino acid substitution introduced into the parent virus was antigenically important it was expected to cause a decrease in HI titer for the mutant virus against antisera induced by parent-like virus (ΔH_1, Eq 5) and a corresponding increase against antisera raised against mutant-like virus (ΔH_2, Eq 5). Conversely, a change in virus receptor-binding avidity is expected to cause a consistent decrease (or increase) in titer with these two groups of antisera (ΔH_1 and ΔH_2). Therefore, for each substitution the associated change in $\log_2$ HI titer, relative to Neth93 or Neth93 Δ130, were partitioned into antigenic (ΔH_A) and non-antigenic (ΔH_N) components.

The predicted antigenic effect of each substitution (from Table 1) is shown alongside mean observed changes in $\log_2$ HI titer partitioned into antigenic (ΔH_A) and non-antigenic effects (ΔH_N) in Table 2. The range of antigenic effects of K141E, ΔK130, E153K and D187N amino acid substitutions, measured against the panel of antisera, were consistent with predictions from the modeling. The range in antigenic impact (ΔH_A) measured using individual antisera is shown in Fig 3 and the mean observed titers averaged across four repeats are shown in S4 Table and as a heat-map in S2 Fig. Across all substitutions we observed a mean error in our predictions of 0.14 antigenic units.

The predicted and observed antigenic impacts, based on HI results, shown in Table 2 and Fig 3 indicate that our model captures the mean impacts of the HA1 amino acid substitutions identified. However, we also observed non-antigenic effects (ΔH_N) of substitutions that resulted in higher or lower HI titers irrespective of antigenic similarity between test virus and the reference virus against which a particular antiserum was raised. Such effects exceeding 0.78 antigenic units, shown in bold in Table 2, cannot be explained solely by differences in virus concentration resulting from the limited accuracy of the hemagglutination assay, used to standardize hemagglutinating units prior to HI. We observed a relatively small antigenic impact of the E153K substitution, but a large non-antigenic effect, on HI titer. This result is supported by previous work showing E153K to have a relatively small impact on monoclonal antibody
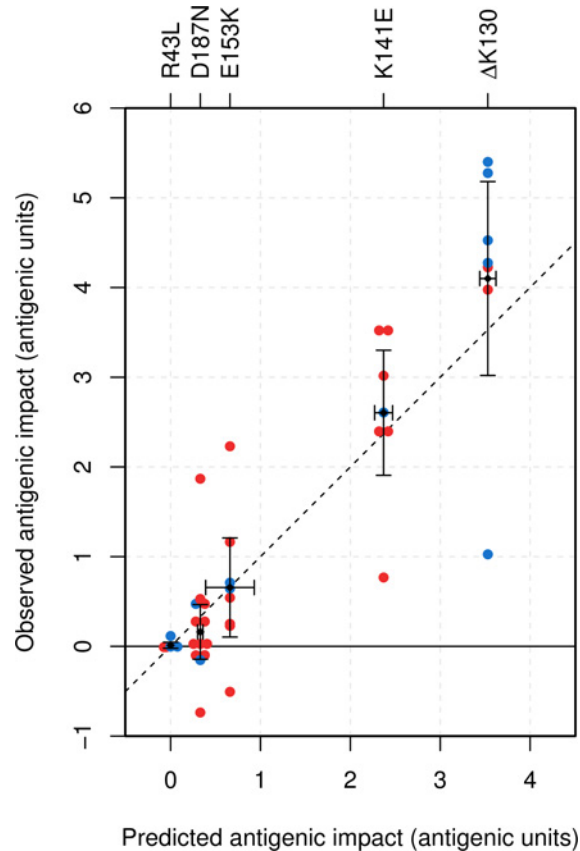
Fig 3. Observed and predicted antigenic impact of amino acid substitutions. The mean antigenic impact of each substitution predicted from modeling (Table 1) plotted against the mean observed impact averaged across antisera in the panel (S3 Table). 95% confidence intervals are shown for both. Each point shows the observed mean antigenic impact (ΔH_A, change in HI titer for a recombinant virus relative to its parent virus) of a particular amino acid substitution (labeled at top) with each antiserum in the panel. Red points indicate that the reference virus lacked the amino acid substitution, so the predicted impact of mutation is a reduction in titer; blue points indicate that the reference virus shared the substitution, so the predicted impact of mutation is an increase in titer. The number of points for each substitution is dependent on whether it was inserted into one or both (Neth93 and Neth93 Δ130) parental viruses and on the number of antisera used. A negative observed antigenic impact indicates a change in HI titer in the opposite direction to that predicted. Mean titers used to calculate antigenic and non-antigenic effects of substitutions are shown in S4 Table and as a heat-map in S2 Fig.

doi:10.1371/journal.ppat.1005526.g003

binding while causing a large increase in receptor-binding avidity that together contributed to large reductions in HI titer indicating escape from inhibition by polyclonal antiserum [27].

Antigenic cartography

Among the substitutions we identified, the mean antigenic effect of two, ΔK130 and K141E, was estimated to be greater than two antigenic units. Each of these substitutions has been previously identified as antigenically important (ΔK130 [25], K141E [12]), and K141E has been associated with a transition between clusters of antigenically distinct viruses in the antigenic evolution of H1N1 [12]. To visualize antigenic evolution, viruses and antisera were positioned in a two dimensional antigenic map using a Bayesian multidimensional scaling model that estimates reference virus immunogenicity and test virus receptor-binding avidity [28]. Examination of the resulting antigenic map showed that the substitutions ΔK130 and K141E could explain the two transitions between antigenic clusters that occurred during the period of H1
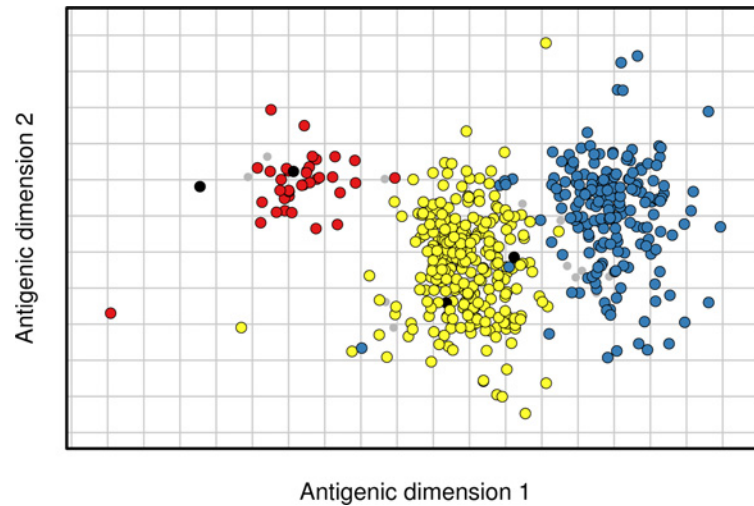
Fig 4. Position of substitutions ΔK130 and K141E on an antigenic map. Map locations are shown for a representative example from a Bayesian multidimensional scaling model that estimates virus location, antiserum location, reference virus immunogenicity and test virus receptor-binding avidity. Gridlines represent single antigenic units, two-fold dilutions in the HI assay. Viruses are shown as colored circles and antisera are shown as grey points. Viruses are colored in relation to the substitutions ΔK130 and K141E: 130K 141K (red, n = 36), Δ130 141K (yellow, n = 273), Δ130 141E (blue, n = 193). Viruses with other amino acid combinations are colored black (n = 4).

evolution studied ([Fig 4](#)), and were the only antigenic substitutions identified that separated these clusters.

Sequence-based prediction

Finally, to assess whether the 18 inferred antigenic determinants shown in [Table 1](#) were predictive of antigenic phenotype, as assessed by HI titer, our results were cross-validated 100 times using model parameters derived from 90% of the pairs of virus and antiserum in the data, randomly selected each time. These 18 included two cases with multiple, co-occurring HA substitutions (R43L, F71I, S271P with ΔK130, and E74G with E120G) and 16 single substitutions. ΔK130 was selected rather than the co-occurring substitutions R43L, F71I or S271P given the results of genetics experiments described in [Table 2](#) and [Fig 3](#), however a single ambiguous term was used to represent either E74G or E120G as these have not been discriminated between.

$$\log_2(H_{r,v}) = k_0 + \sum\nolimits_{\omega \in \Omega} m_\omega \alpha_\omega(r, v) + s_r + a_v \tag{6}$$

[Eq 6](#) describes the predictive model, based on [Eq 1](#), that estimates the antigenic dissimilarity of two viruses based on which substitutions separate them. α_ω is 1 when reference virus (r) and test virus (v) are separated by a specific substitution (or its reverse), and 0 otherwise. The substitutions included in the model are Ω = {S36N, S72F, E74G or E120G, ΔK130, K141E, S142N, E153G, E153K, G153K, K163N, S183P, N184S, D187N, D187V, A190T, W252R, E274K, R313K}—all of the substitutions identified above. Each substitution in Ω also has an associated antigenic impact, m_ω, previously identified in [Table 1](#), but here estimated repeatedly in the model from (90% of the pairs of virus and antiserum) training data to predict the (10% of the pairs of virus and antiserum) test data in a cross-validation procedure. We compared the prediction error of this model with and without the parameters a_v and s_r to investigate the importance of including non-antigenic effects in the model, and also with the subset of Ω containing
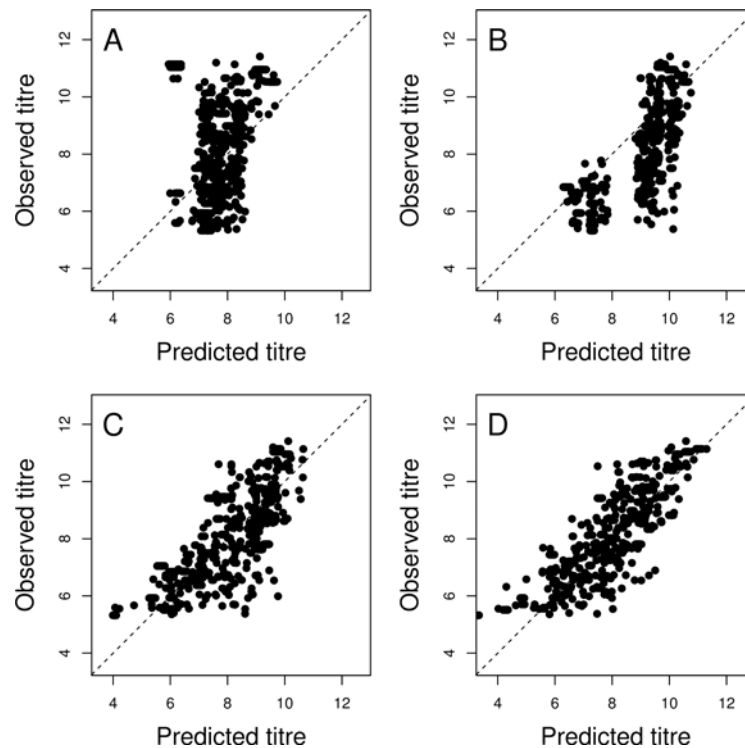
Fig 5. Sequence-based prediction of antigenic phenotype. Observed and predicted HI titers plotted on log$_2$ scale (antigenic units) using representative models trained with data for 90% of the pairs of virus and antiserum. Predictive models contained terms for A) Average titers for each reference virus, B) Antigenic cluster-defining substitutions ΔK130 and K141E, C) All 18 antigenic substitution(s) shown in Table 1, D) All 18 antigenic substitution(s) shown in Table 1 with additional terms that estimate differences in test virus receptor-binding avidity (non-antigenic variation in titer associated with each virus). Each model was fitted to the same training dataset comprising 90% of all pairs of virus and antiserum and predictions for the remaining data are shown. Incremental improvements in mean absolute prediction error are shown alongside SEM and 95% upper limits in S5 Table.

doi:10.1371/journal.ppat.1005526.g005

only substitutions that defined the clusters found on the antigenic map ($\Omega' = \{\Delta K130, K141E\}$) to investigate the importance of low-impact substitutions.

A simple null model that contained no substitution terms and that used only the average titer for antiserum raised against each reference virus (s_r) to predict antigenic phenotype, produced a mean absolute error of 1.44 units (Fig 5A). By including only identified cluster-defining substitutions during the period studied, ΔK130 and K141E, prediction improved, with a mean absolute error of 1.06 antigenic units (Fig 5B). Adding in all 18 substitution(s) (Table 1) reduced this to an error of 0.75 units (Fig 5C). Inclusion of lower-impact, non-cluster defining substitutions therefore allowed more accurate prediction of antigenic phenotype. Prediction accuracy was further improved by compensating for non-antigenic differences between viruses (*i.e.* variation in their receptor-binding avidity). Allowing the average titer for each virus to absorb this variation and using a_v terms for prediction resulted in an error of only 0.65 antigenic units (Fig 5D). When a virus that is not present in the training data appears in the test data the reactivity parameter (a_v) associated with that virus is set to the mean of the training virus reactivities. We calculate the mean absolute residual error of the assay itself, after controlling for day-to-day variability, from the estimated underlying average titers (Eq 7) of the frequently (≥ 5) observed pairs as 0.47 antigenic units, with a lower 95% credible interval of 0–1.21. This provides a lower bound below which we would be overfitting the unrepeated (non-reference virus) titres. This is partly a result of the discrete nature of the assay results, and partly the inherent variability
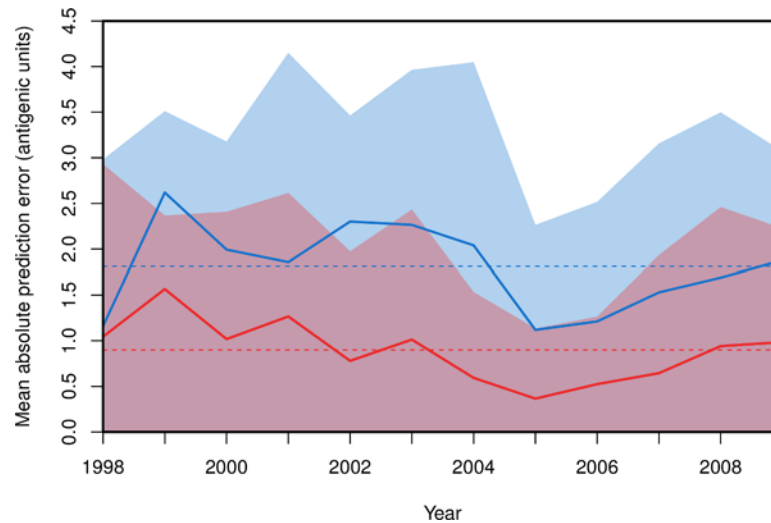
Fig 6. Prediction error through time for models used to predict HI titers of viruses isolated in the following year. The mean, absolute difference between observed titers for viruses isolated in a given year and titers predicted using models trained to HI data collected in previous years is shown. Predictive models included terms for cluster-defining substitutions ΔK130 and K141E only (solid blue line) or for all 18 substitutions in Table 1 (solid red line). For each model, shaded areas show the lower 95% credible interval on the absolute prediction error. In each year the blue 95% credible interval extends vertically on the y-axis above the red 95% credible interval. Mean, absolute prediction errors averaged across the twelve years are shown as dashed lines.

doi:10.1371/journal.ppat.1005526.g006

of the assay. This error is therefore the minimum threshold against which our errors should be measured, giving an additional error of 0.18 antigenic units for the full model.

However, autocorrelation between training and test datasets is still present here, since the same viruses are present in both datasets. To further investigate the predictive power of the approach while reducing the impact of autocorrelation, the antigenic phenotype of viruses isolated in each year of our dataset (from 1998 to 2009) was predicted using model parameters derived using only data collected prior to that year. We modeled the HI titers of each test virus isolated in a year using antisera raised against viruses isolated in all previous years. Mean absolute prediction error averaged across the twelve years was 1.81 units (SEM = 0.0112, SD = 0.92) when only the substitutions ΔK130 and K141E were included. This reduced to 0.90 units (SEM = 0.0085, SD = 0.70) when all 18 substitutions were included. The mean absolute prediction error of each model in each year is shown in Fig 6, and the fit of the predicted to the observed titres over all of the years is shown in S3 Fig in a variety of models; in every year the accuracy is improved by the inclusion of the lower-impact antigenic determinants in addition to the two substitutions that define antigenic clusters on the map. Variation in titre that is within one well in HI (1 antigenic unit) is treated by the WHO Global Influenza Surveillance and Response System (GISRS) as antigenically negligible, and for high-growth reassortant viruses to be considered antigenically parent-like, viruses must be recognized by antisera raised against the reassortant and prototype/parent at titres less than 4-fold (2 antigenic units) of the homologous viruses in HI assays [21]. This suggests that our full model is making predictions that are of a useful level of accuracy.

Discussion

Using a modeling approach that integrated HA sequence data and HI antigenic data for over 500 viruses, we have identified substitutions responsible for the antigenic evolution of former

seasonal influenza A(H1N1) viruses over a period of more than 10 years. We identified 18 sub-
stitutions at 15 amino acid positions in HA1: two that had high-impact on antigenicity (which
individually can lead to a need to change vaccine virus), and 13 of lower impact, including
some too low to be directly observable in routine HI assays. Substitutions at four of the fifteen
amino acid positions occurred multiple times in the evolutionary history of the virus, consis-
tently explaining observed antigenic changes. Antigenic cartography of the H1N1 viruses iden-
tified only three antigenic clusters, transitions between which can be explained by the two
high-impact substitutions. Our more detailed findings could support our knowledge of the
genetic basis of antigenic variation of the currently circulating A(H1N1)pdm09 viruses.
Though these viruses have exhibited little antigenic variability so far, viruses with substitutions
in the 153–157 region, including K153E which we specifically identify here (K154E in cited
work), of the HA showing reduced reactivity to ferret antisera raised against the A/California/
7/2009 vaccine virus have been noted [2,21]. Subsequent reverse genetics experiments have
identified escape mutants possessing substitutions at position 153 including substitutions we
detect here (K153E and K153G) [29].

As here with H1N1, in antigenic maps of H3N2 viruses antigenic distances between viruses
belonging to the same antigenic cluster often exceeded distances between viruses in adjacent
clusters, demonstrating the need to assess non-cluster defining substitutions [11]. Further, the
selection of approximately twice as many H3N2 vaccine viruses as antigenic clusters identified
by Smith *et al*. during the period 1968 to 2003 is supportive of the importance of substitutions
not readily identified using antigenic maps [12]. Using cross-validation we show that inclusion
of substitutions causing low- to high-impact antigenic changes significantly improved predic-
tion for the viruses of the H1N1 subtype. Furthermore, when predicting the antigenic pheno-
type of viruses in the following year, including these substitutions doubled accuracy; in
contrast, using only cluster-defining substitutions generated worse predictions in each year
and in several years mean absolute prediction error for these models exceeded two antigenic
units. An antigenic distance of two units may necessitate a change of vaccine virus recommen-
dation, so this improvement has significant implications for the usefulness of these predictive
models. The improved accuracy of predictions made using the full model shows that substitu-
tions causing low-impact antigenic changes are common in the evolution of influenza A viruses
and crucial to the tracking of antigenic evolution. It is notable that the model does not consis-
tently improve over time. However, we should not expect this since antigenic novelty continues
to arise, and increased error will always occur in years where important substitutions, not
observed in previous years, are prominent. Identification of substitutions responsible for such
smaller incremental changes in antigenicity also raises the prospect of fine-tuning vaccine
viruses by mutating existing candidate vaccine viruses or their derivatives.

The model also partitions the results of the HI assay into antigenic and non-antigenic
effects, and we have done the same when characterizing the mutant viruses generated by
reverse genetics. At position 153 we detect a non-antigenic effect of substitution contributing
to apparent antigenic effects in HI titers, consistent with previous studies of former seasonal
H1N1 [27]. Substitutions in the 150-loop (153–157) of HA1 have been shown to occur during
culture of A(H1N1)pdm09 viruses [2,21] and G155E substitution has been shown to affect
receptor-binding specificity or avidity [30]: such receptor-binding alterations probably contrib-
ute to apparent antigenic effects attributed to substitutions introduced into this region of the A
(H1N1)pdm09 HA1 by reverse genetics [29]. Understanding the genetic variation underlying
changes in the receptor-binding avidity of influenza viruses that contribute to apparent anti-
genic effects as measured by HI assay is clearly a very important area for further investigation.
We have also not explicitly modeled how the antigenic impact of substitution at one position
depends on substitutions present at other HA, or neuraminidase, positions. It has been

proposed that epistasis is prevalent in the evolution of influenza surface glycoproteins, however few examples have been confirmed using phenotypic data [31]. In the assessment of viruses derived by reverse genetics performed here, the effect of substitutions varied with the antiserum used in HI assays. Understanding how variation in the antigenic impact of a particular substitution can be attributed to existing genetic and antigenic differences between viruses is another target for future research that should further improve our understanding of how specific substitutions affect antigenicity. These approaches are also applicable to foot-and-mouth disease virus, which we have studied previously [15]. The smaller number of reference viruses and much reduced datasets available there reduce the power of the approach, but we are currently transferring the lessons learnt from the larger influenza datasets to this virus.

Łuksza and Lässig recently demonstrated that they can predict the evolutionary success of influenza clades using population genetics models that incorporate information on the frequencies of genotypes in the previous season and the number of substitutions in known antigenic and non-antigenic regions [32]. However, the effectiveness of such an approach depends on the ability to quantify how individual mutations affect fitness [33]. Antigenicity is a key component of the fitness of human influenza viruses, and we show that we can both quantify the impact of specific amino acid changes and use this knowledge to predict antigenic phenotype directly from sequence data. Our ability to quantify heterogeneities in the antigenic impact of substitutions improves our understanding of the genetic basis of antigenic evolution in influenza viruses. This allows sequence-based predictions of antigenicity to be made for genetic variants before HI testing, thereby increasing the value of sequence data from emerging genetic variants to assist targeting of antigenic analyses and hasten the identification of emergent antigenic variants. We anticipate that the ability to determine roles of both high- and low-impact amino acid substitutions in antigenic drift will complement existing methods and improve genotype-based predictions of virus fitness and consequent evolutionary success.

Materials and Methods

Ethics statement

Human specimens. 1) WHO National Influenza Centres (NICs) are national institutions designated by national Ministries of Health who collect appropriate clinical specimens from patients and undertake virus surveillance and virus identification. The NICs send representative virus isolates to a WHO Collaborating Centre for Reference and Research on Influenza (WHO CCs). Many thousands of these specimens are received each year by the WHO CCs. The identifies of the samples collected are anonymized prior to sharing.

Madin-Darby Canine Kidney Epithelial cells. MDCK cells were purchased from the American Type Culture Collection (ATCC) and maintained at Mill Hill and Edinburgh laboratories.

Ferrets. Ferrets were from a Home Office approved supplier and housed in containment level 2 in the UK Home Office approved facilities at the Mill Hill laboratories. The studies were approved by the Ethical Review Bodies of The Francis Crick Institute and the MRC National Institute for Medical Research and licensed by the UK Home Office under license numbers 80/2541, 80/2090 and previous licenses under the UK Animals (Scientific Procedures) Act 1986. Infected ferrets were monitored closely with respect to their health (e.g. lethargy, inability to feed/drink, etc.) and any that considered to be showing severe symptoms were culled by terminal anesthesia. Two weeks post-infection the ferrets were put under terminal anesthesia using a specific mixture of drugs (Vetalar, Rompun and Pentoject) dependent on the weight of the ferret and were exsanguinated to provide antiserum for use in HI studies.

Data

Viruses were originally isolated from clinical specimens either by WHO NICs or by the WHO Collaborating Centre. The antigenic dataset encompassed 506 former seasonal A(H1N1) viruses for which HA gene sequence data were available. Forty-three of these 506 viruses were chosen as reference viruses and these were used to generate antiserum for use in HI studies. All HI data used were obtained using post-infection ferret antisera. In total, 19,905 HI titers measured between 3,734 unique combinations of virus and antiserum, made on 351 dates from 1997 to 2009 were analyzed. The data associated with this study are available online [34].

Recombinant viruses

Viruses were generated using a protocol based on Hoffman *et al.* 2000 [35]. HA and neuraminidase cDNAs of A/Netherlands/1/93 (Neth93), which had been exclusively propagated in cell culture, were amplified using a standard RT-PCR protocol. These cDNAs were cloned into the pHW2000 vector. Mutations were introduced into the HA plasmid using the QuikChange lightning site-directed mutagenesis kit (Agilent Technologies, Santa Clara, California). Co-cultured 293T and MDCK cells were co-transfected with plasmids containing HA and neuraminidase derived from Neth93 with the remaining six genes from A/Puerto Rico/8/34. After 2–3 days, recombinant viruses in the supernatant of transfected cells were recovered and propagated in MDCK cells as described in Lin *et al.* [36]. Virus HA sequences were verified after passage.

HI assays and analysis

HI assays were performed on recombinant viruses by standard methods [10]. Post-infection ferret antisera raised against the following reference viruses were used: A/Bayern/7/95, A/Johannesburg/82/96, A/Johannesburg/159/97, A/Ulan-Ude/209/98, A/Hong Kong/4847/98, A/New Caledonia/20/99 and A/Hong Kong/1252/2000. Amino acid identities at HA1 positions 43, 130, 141, 153 and 187 of these seven reference viruses are shown in S3 Table.

Average changes in $\log_2$ HI titer between parent and mutant recombinant viruses were quantified and partitioned into antigenic (ΔH_A) and non-antigenic (ΔH_N) effects using Eq 5. Antigenic effects were compared with predictions from modeling. Mean error in predictions across all substitutions was calculated as the average difference between the predicted mean and each measured antigenic change in HI using a specific virus dilution measured against a particular antiserum (excluding measurements restricted by the lower threshold of the HI assay).

Small non-antigenic changes in HI titer (ΔH_N) between two viruses could be explained by the routine standardization of both viruses using the hemagglutination assay prior to HI. Limitations in the accuracy of the hemagglutination assay controlling for virus concentration (± 0.5 hemagglutinating units) for both parent and mutant viruses mean that effects on HI titer below 0.78 antigenic units (95% CI) could be a result of test error. Corresponding antigenic changes, however, look at differences between antisera for a single sample of the same diluted virus, controlling for this effect.

Phylogenetic analysis

HA1 nucleotide sequences of the 506 viruses were aligned using MUSCLE [37]. Phylogeny construction and analysis was carried out using BEAST v1.7.4 [22] which uses Markov chain Monte Carlo (MCMC) to explore parameter space and evaluate phylogenetic models and Tracer v1.5 [38]. Phylogenies were estimated using a variety of nucleotide substitution and

molecular clock models. A relaxed, uncorrelated clock and a GTR+I+Γ_4 nucleotide substitution model were determined to be most suitable through comparison of Bayes factors [39]. Bayes factor analysis also determined that a separate partition should be created for the third codon position to allow rates of nucleotide substitution at this position to vary relative to the first and second codon positions. As a prior, we assumed an underlying coalescent process with a constant population size on the tree. Time of isolation was used to calibrate the molecular clock allowing rates of evolution along branches to be estimated. The maximum clade credibility tree was identified from a posterior sample of 10,000 trees. Substitution at position 187, associated with adaptation to propagation in eggs [40–42], was assumed to be an artifact with potential to distort phylogenetic inference, so nucleotides coding for position 187 were excluded from phylogenetic analysis. Ancestral amino acid state at each node in the phylogeny for each position identified by modeling was estimated using the FLU amino acid substitution model [43] and unlinked strict molecular clocks for each amino acid position.

Antigenic cartography

Virus locations in antigenic space were estimated using the Bayesian multidimensional scaling technique of Bedford *et al.* [28], which extends Smith *et al.* [11] by incorporating a phylogenetic diffusion process and estimates of antiserum and virus reactivity to account for variation in the immunogenicity of different reference viruses and in the receptor-binding avidity of viruses.

Mixed effects modeling and model selection

Co-variance between HI titer and the size of residuals from models using linear HI titers necessitated the use of logarithmically transformed HI titer as the response variable (to ensure homoscedasticity). Base 2 was chosen (without loss of generality) for the logarithm to follow Smith *et al.* and work throughout was in terms of $\log_2$ (or *antigenic units*) where 1 corresponds to a two-fold dilution of antiserum in the HI assay [11]. The likelihood ratio test was used to test models containing combinations of the following explanatory variables: the reference virus against which the antiserum was raised, the test virus, and the date on which the assay was performed.

Following Reeve et al. [15], each branch of the HA1 phylogeny was tested in the model as a fixed effect term. Each branch term was included as a discrete indicator variable: 1 when reference virus and test virus were separated by the branch in the phylogenetic tree and 0 otherwise. Random restart hill-climbing was used to determine the best model [44]. To a random consistent starting model, branch terms were added and removed at random to maximize model fit, assessed by AIC [45]. This was repeated while randomizing their order to identify the best model to avoid sensitivity to the order in which the parameters were presented. This approach was conservative since it was used to determine the branches used to control for phylogenetic correlations in the data, and adding in extra unnecessary terms simply reduced the power of the analysis. We show elsewhere that sparse Bayesian variable selection methods are more powerful for such problems, but are not computationally feasible for use on such large datasets [46]. Conceptually similar machine learning techniques have been used on related influenza datasets [19], but these did not control for phylogenetic correlation, which we consider to be critical to avoid false positives, and this remains the computationally expensive step.

PyMOL Molecular Graphics System v1.7.7.2 (http://www.pymol.org) was used to visualize and identify positions exposed on the external surface of the HA 3D-structure of A/Puerto Rico/8/34 (resolved to 2.3Å, Protein Data Bank ID: 1RU7) [24], which had previously been identified as having significant solvent accessibility using naccess v.2.1.1 [47] (solvent

accessibility over 18 Å^2). To account for potential structural changes during the period of evolution since the isolation of A/Puerto Rico/8/34, the HA 3D-structure of A/Solomon Islands/3/2006 (resolved to 3.19Å, Protein Data Bank ID: 3SM5) [48] was also used. Amino acid dissimilarity between reference virus and test virus at each position exposed on the surface of either structure, not conserved within the dataset, was tested as a predictor of reduced HI titer (p<0.05) using a Holm-Bonferroni correction to account for multiple tests [23]. Although position 187 was excluded from phylogenetic inference, amino acid dissimilarity at this position was tested as a predictor of antigenic difference. At each HA position identified at this stage, the mean antigenic impact of specific amino acid substitutions was determined by examining the associated parameter (k_1, Eq 3) using data subsets with only two amino acid variants at that position.

Amino acid positions at which substitution correlated with antigenic change at only a single position in the phylogeny were identified next. Branches retained by the random restart hill-climbing approach, but not correlated with any substitutions that explained antigenic change in multiple branches, were identified by including terms for those substitutions in the model, and examining the resultant regression parameter (m_i, Eq 4) associated with each branch. Branch terms whose effects remained detectable ($m_i > 0.25$) were removed sequentially. Amino acid dissimilarity between reference virus and test virus at each non-conserved surface-exposed position was re-tested for inclusion in the model in the absence of each branch term. Statistical analyses were performed using R software [49] and the package lme4 [50].

Sequence-based prediction

To test the predictive power of the identified substitutions, repeated randomized cross-validation was used: 90% of pairs of virus and antiserum were repeatedly selected at random (100 times) to act as the training dataset to which models were fitted allowing predictions of titer to be made for the test dataset composed of the remaining 10% pairs. Branch terms included in the antigenic site analysis, to prevent identifications of false positives, were not included in these predictive models since it is only the substitutions themselves that are causally connected to antigenic change.

HI titers are affected by a number of experimental variables and thus observed titers are affected by substantial experimental variability as illustrated in S1 Fig. Therefore, we compared predictions with the estimated underlying average titers between antiserum (r) and test virus (v), fitted using the linear mixed-effects model that best fitted the data (Eq 7). In addition to fixed effects for antiserum (s_r), test virus reactivity (a_v) and the interaction between them (γ_{rv}, which estimates their antigenic relationship directly), ε_D and ε_R encapsulate a random effect for date and residual measurement error, respectively.

$$\log_2(H_{r,v}) = k_0 + s_r + a_v + \gamma_{rv} + \varepsilon_D + \varepsilon_R \tag{7}$$

Additional to using 10% of the data as a test dataset, we examined the ability of the same models to predict antigenic relationships between existing antisera and 'future' viruses. Test datasets containing all observations between viruses isolated in a given year and antisera raised against reference viruses isolated in all previous years were constructed. Data from all previous years were used as training datasets. Models were implemented using JAGS v3.3.0 [51] through R using the package runjags [52].

Supporting Information

S1 Fig. Distributions of observed HI titers for the most frequently used antigen within the dataset. Frequency of HI titers recorded for the virus A/New Caledonia/20/99 tested using

antisera raised against A/New Caledonia/20/99 (**A: Homologous**) and A/Beijing/262/95 (**B: Heterologous**).
(TIF)

S2 Fig. Heat-map and clustering analysis of HI titers with mutant viruses. Hierarchical clustering of the wild type Neth93 and mutant viruses generated from it by reverse genetics. Reference strains used to generate antisera are arranged according to their ability to inhibit agglutination of turkey RBCs by each virus. Viruses are simultaneously clustered along the vertical axis according to their antigenic profile. Dendograms indicating antigenic relatedness are shown at the top (for antisera) and to the left (for viruses). Coloring represents $\log_2$ HI titer as indicated at top left with the histogram showing the frequency (count) for each titer.
(TIF)

S3 Fig. Sequence-based prediction of antigenic phenotype of viruses isolated in years following period of collection of data used to train model. Observed and predicted HI titers plotted on $\log_2$ scale (antigenic units) using models trained to data collected prior to a given year and used to predict titres for viruses isolated in that given year. Plots show predictions aggregated across 12 years (1998–2009). Predictive models contained terms for A) The estimated baseline titre B) Average titers for each reference virus, C) Antigenic cluster-defining substitutions ΔK130 and K141E, D) All 18 antigenic substitution(s) shown in Table 1.
(TIF)

S1 Table. Model quality as assessed by AIC. Models are shown alongside their Δ AIC, the improvement in AIC relative to a null, intercept model (model A). Model terms are consistent with notation used in Eqs 1–7 and are described in full in S2 Table. Models A-H include every combination of base terms introduced in Eq 1. Relative AIC scores led to H being strongly preferred and other combinations were therefore discounted. Models I and J contain 62 branch terms identified using Eq 2. Model J also includes terms for seven substitutions identified using Eq 3. Model K generates a fitted value for every observed combination of reference virus and test virus and corresponds to Eq 7. Models L and M include only substitution terms to explain antigenic differences and no branch terms. Model L includes all 18 identified substitutions (7 identified using Eq 3 and 11 identified using Eq 4). Model M contains terms for the substitutions of highest antigenic impact (K141E and ΔK130) only.
(DOCX)

S2 Table. Reference guide for terms used in S1 Table and in equations throughout the text.
(DOCX)

S3 Table. Antisera used to characterize recombinant viruses. Amino acid identity at HA positions 43, 130, 141, 153 and 187 of reference viruses against which antisera were raised and used to antigenically characterize recombinant viruses (* indicates deletion of amino acid corresponding to position 130). Cells are colored according to whether the reference virus against which antiserum was raised lacked or shared each amino acid substitution introduced by mutagenesis to produce recombinant viruses (R43L, ΔK130, K141E, E153K and D187N). Red indicates that the reference virus lacked the substitution introduced into the recombinant virus and so was in the ancestral state (e.g. R43) and blue indicates that the reference virus shared the introduced substitution (e.g. L43). Absence of color indicates that the reference amino acid identity at the position of substitution in the recombinant virus was different from both of the parental viruses (Neth93 and Neth93 Δ130) and from the mutant virus.
(DOCX)

S4 Table. Mean HI titers for recombinant viruses measured against antisera presented in S3 Table. Geometric mean HI titers are recorded as the reciprocal of the highest dilution of a particular antiserum that inhibited hemagglutination of a standardized concentration of red blood cells by eight hemagglutinating units of each recombinant virus. A visual description of these data is provided in S2 Fig.
(DOCX)

S5 Table. Average absolute prediction error (antigenic units) across various predictive models. * Predictions for this model were made using a parameter for date of test. Date was not used as a predictor in any other model. † This is the upper limit on the lower 95% credible interval for the mean, absolute test error.
(DOCX)

Acknowledgments

Author Contributions

Conceived and designed the experiments: WTH DJB RSD DTH AJH JWM RR. Performed the experiments: WTH DJB JPJH VG. Analyzed the data: WTH RR. Contributed reagents/materials/analysis tools: DJB TB. Wrote the paper: WTH RSD TB AJH JWM RR.

References

1. WHO (2009) WHO Influenza fact sheet.

2. Klimov AI, Garten R, Russell C, Barr IG, Besselaar TG, et al. (2012) WHO recommendations 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 2011. Vaccine 30: 6461–6471. doi: 10.1016/j.vaccine.2012.07.089 PMID: 22917957

3. Skehel JJ, Wiley DC (2000) Receptor binding and membrane fusion in virus entry: The influenza hemagglutinin. Annu Rev Biochem 69: 531–569. PMID: 10966468

4. Wiley DC, Wilson IA, Skehel JJ (1981) Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. Nature 289: 373–378. PMID: 6162101

5. Caton AJ, Brownlee GG, Yewdell JW, Gerhard W (1982) The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 subtype). Cell 31: 417–427. PMID: 6186384

6. Webster RG, Laver WG, Air GM, Schild GC (1982) Molecular mechanisms of variation in influenza viruses. Nature 296: 115–121. doi: 10.1038/296115a0 PMID: 6174870

7. 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 U S A 94: 7712–7718. PMID: 9223253

8. Nelson MI, Holmes EC (2007) The evolution of epidemic influenza. Nat Rev Genet 8: 196–205. doi: 10.1038/nrg2053 PMID: 17262054

9. Hirst GK (1942) The quantitative determination of influenza virus and antibodies by means of red cell agglutination. J Exp Med 75: 49–64. PMID: 19871167

10. WHO (2011) Manual for the Laboratory Diagnosis and Virological Surveillance of Influenza. http://whqlibdoc.who.int/publications/2011/9789241548090_eng.pdf.

11. 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. doi: 10.1126/science.1097211 PMID: 15218094

12. Koel BF, Burke DF, Bestebroer TM, van der Vliet S, Zondag GCM, et al. (2013) Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution. Science 342: 976–979. doi: 10.1126/science.1244730 PMID: 24264991

13. Bedford T, Cobey S, Pascual M (2011) Strength and tempo of selection revealed in viral gene genealogies. BMC Evol Biol 11: e220. doi: 10.1186/1471-2148-11-220

14. Neher RA, Russell CA, Shraiman BI (2014) Predicting evolution from the shape of genealogical trees. Elife 3: e03568. doi: 10.7554/eLife.03568

15. Reeve R, Blignaut B, Esterhuysen JJ, Opperman P, Matthews L, et al. (2010) Sequence-based prediction for vaccine strain selection and identification of antigenic variability in foot-and-mouth disease virus. PLoS Comput Biol 6: e1001027. doi: 10.1371/journal.pcbi.1001027 PMID: 21151576

16. Lee M-S, Chen JS-E (2004) Predicting antigenic variants of influenza A/H3N2 viruses. Emerg Infect Dis 10: 1385–1390. PMID: 15496238

17. Huang J-W, Lin W-F, Yang J-M (2012) Antigenic sites of H1N1 influenza virus hemagglutinin revealed by natural isolates and inhibition assays. Vaccine 30: 6327–6337. doi: 10.1016/j.vaccine.2012.07.079 PMID: 22885274

18. Steinbrück L, McHardy AC (2012) Inference of genotype-phenotype relationships in the antigenic evolution of human influenza A (H3N2) viruses. PLoS Comput Biol 8: e1002492. doi: 10.1371/journal.pcbi.1002492 PMID: 22532796

19. Sun H, Yang J, Zhang T, Long L, Jia K, et al. (2013) Using sequence data to infer the antigenicity of influenza virus. MBio 4: e00230–13. doi: 10.1128/mBio.00230-13 Updated. PMID: 23820391

20. Lin YP, Gregory V, Collins P, Kloess J, Wharton S, et al. (2010) Neuraminidase receptor binding variants of human influenza A(H3N2) viruses resulting from substitution of aspartic acid 151 in the catalytic site: A role in virus attachment? J Virol 84: 6769–6781. doi: 10.1128/JVI.00458-10 PMID: 20410266

21. Barr IG, Russell C, Besselaar TG, Cox NJ, Daniels RS, et al. (2014) WHO recommendations for the viruses used in the 2013–2014 Northern Hemisphere influenza vaccine: Epidemiology, antigenic and genetic characteristics of influenza A(H1N1)pdm09, A(H3N2) and B influenza viruses collected from October 2012 to January 2013. Vaccine 32: 4713–4725. doi: 10.1016/j.vaccine.2014.02.014 PMID: 24582632

22. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29: 1969–1973. doi: 10.1093/molbev/mss075 PMID: 22367748

23. Holm S (1979) A simple sequentially rejective multiple test procedure. Scand J Stat 6: 65–70.

24. Gamblin SJ, Haire LF, Russell RJ, Stevens DJ, Xiao B, et al. (2004) The structure and receptor binding properties of the 1918 influenza hemagglutinin. Science 303: 1838–1842. doi: 10.1126/science.1093155 PMID: 14764886

25. 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. J Gen Virol 88: 3209–3213. doi: 10.1099/vir.0.83184–0 PMID: 18024888

26. Daniels RS, Douglas A, Skehel JJ, Wiley D, Naeve W, et al. (1984) Antigenic analyses of influenza virus haemagglutinins with different receptor-binding specificities. Virology 138: 174–177. PMID: 6208680

27. Hensley SE, Das SR, Bailey AL, Schmidt LM, Hickman HD, et al. (2009) Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift. Science 326: 734–736. doi: 10.1126/science.1178258 PMID: 19900932

28. Bedford T, Suchard MA, Lemey P, Dudas G, Gregory V, et al. (2014) Integrating influenza antigenic dynamics with molecular evolution. Elife 3: e01914. doi: 10.7554/eLife.01914 PMID: 24497547

29. Koel BF, Mögling R, Chutinimitkul S, Fraaij PL, Burke DF, et al. (2015) Identification of amino acid substitutions supporting antigenic change of influenza A(H1N1)pdm09 viruses. J Virol 89: 3763–3775. doi: 10.1128/JVI.02962-14 PMID: 25609810

30. Liu Y, Childs RA, Matrosovich T, Wharton S, Palma AS, et al. (2010) Altered receptor specificity and cell tropism of D222G hemagglutinin mutants isolated from fatal cases of pandemic A(H1N1) 2009 influenza virus. J Virol 84: 12069–12074. doi: 10.1128/JVI.01639-10 PMID: 20826688

31. Kryazhimskiy S, Dushoff J, Bazykin GA, Plotkin JB (2011) Prevalence of epistasis in the evolution of influenza A surface proteins. PLoS Genet 7: e1001301. doi: 10.1371/journal.pgen.1001301 PMID: 21390205

32. Łuksza M, Lässig M (2014) A predictive fitness model for influenza. Nature 507: 57–61. doi: 10.1038/nature13087 PMID: 24572367

33. Koelle K, Rasmussen DA (2014) Prediction is worth a shot. Nature 507: 47–48. doi: 10.1038/nature13054 PMID: 24572355

34. Gregory V, Harvey WT, Daniels RS, Reeve R, Whittaker L, et al. (2016) Human former seasonal Influenza A(H1N1) haemagglutination inhibition data from 1978–2009 from the WHO Collaborating Centre for Reference and Research on Influenza, London, UK. The Crick Worldwide Influenza Centre, The Francis Crick Institute, Mill Hill Laboratory, The Ridgeway, Mill Hill, London NW7 1AA, UK. doi: 10.5525/gla.reasearchdata.289

35. Hoffmann E, Neumann G, Kawaoka Y, Hobom G, Webster RG (2000) A DNA transfection system for generation of influenza A virus from eight plasmids. Proc Natl Acad Sci U S A 97: 6108–6113. doi: 10.1073/pnas.100133697 PMID: 10801978

36. Lin YP, Xiong X, Wharton SA, Martin SR, Coombs PJ, et al. (2012) Evolution of the receptor binding properties of the influenza A (H3N2) hemagglutinin. Proc Natl Acad Sci U S A 109: 21474–21479. doi: 10.1073/pnas.1218841110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1218841110 PMID: 23236176

37. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797. doi: 10.1093/nar/gkh340 PMID: 15034147

38. Rambaut A, Drummond AJ (2009) Tracer v1.5 http://tree.bio.ed.ac.uk/software/tracer/.

39. Suchard MA, Weiss RE, Sinsheimer JS (2001) Bayesian selection of continuous-time Markov chain evolutionary models. Mol Biol Evol 18: 1001–1013. PMID: 11371589

40. Raymond FL, Caton AJ, Cox NJ, Kendal AP, Brownlee GG (1986) The antigenicity and evolution of influenza H1 haemagglutinin, from 1950–1957 and 1977–1983: Two pathways from one gene. Virology 148: 275–287. PMID: 3942036

41. Robertson JS, Bootman JS, Newman R, Oxford JS, Daniels RS, et al. (1987) Structural changes in the haemagglutinin which accompany egg adaptation of an influenza A(H1N1) virus. Virology 160: 31–37. PMID: 3629978

42. Gambaryan AS, Robertson JS, Matrosovich MN (1999) Effects of egg-adaptation on the receptor-binding properties of human influenza A and B viruses. Virology 258: 232–239. doi: 10.1006/viro.1999.9732 PMID: 10366560

43. Dang CC, Le QS, Gascuel O, Le VS (2010) FLU, an amino acid substitution model for influenza proteins. BMC Evol Biol 10: e99. doi: 10.1186/1471-2148-10-99

44. Russell S, Norvig P (1995) Artificial intelligence: A modern approach. Artificial Intelligence: A Modern Approach. Pearson Education Limited.

45. Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Contr 19: 716–723.

46. Davies V, Reeve R, Harvey W, Maree FF, Husmeier D (2014) Sparse Bayesian variable selection for the identification of antigenic variability in the foot-and-mouth disease virus. J Mach Learn Work Conf Proc 33: 149–158.

47. Hubbard S, Thornton J (1993) NACCESS, computer program. Department of Biochemistry Molecular Biology, University College London (http://www.bioinf.manchester.ac.uk/naccess/).

48. Whittle JRR, Zhang R, Khurana S, King LR, Manischewitz J, et al. (2011) Broadly neutralizing human antibody that recognizes the receptor-binding pocket of influenza virus hemagglutinin. Proc Natl Acad Sci U S A 108: 14216–14221. doi: 10.1073/pnas.1111497108 PMID: 21825125

49. R Core Team (2015) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

50. Bates D, Maechler M, Bolker B (2012) Linear mixed-effects models using S4 classes. http://lme4.r-forge.r-project.org/.

51. Plummer M (2012) Just Another Gibbs Sampler v3.3.0 (JAGS): A program for analysis of Bayesian graphical models using Gibbs sampling. http://mcmc-jags.sourceforge.net.

52. Denwood MJ (2013) runjags: Interface utilities for MCMC models in Just Another Gibbs Sampler (JAGS) using parallel and distributed computing methods. http://cran.r-project.org/web/packages/runjags/index.html.