Prediction, dynamics, and visualization of antigenic phenotypes of seasonal influenza viruses

Richard A. Neher,[1] Trevor Bedford,[2] Rodney S. Daniels,[3,4] Colin A. Russell,[5] and Boris I. Shraiman[6]

[1] *Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany*
[2] *Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA*
[3] *The Francis Crick Institute, Worldwide Influenza Centre (WIC),*
Mill Hill Laboratory, The Ridgeway, London NW7 1AA, UK
[4] *formerly the WHO CC for Reference and Research on Influenza, MRC, National Institute for Medical Research, The Ridgeway, London NW7 1AA, UK*
[5] *Department of Veterinary Medicine, University of Cambridge, Cambridge CB3 0ES, UK*
[6] *Kavli Institute for Theoretical Physics, University of California, Santa Barbara, CA 93106, USA*

(Dated: October 6, 2015)

Human seasonal influenza viruses evolve rapidly, enabling the virus population to evade immunity and re-infect previously infected individuals. Antigenic properties are largely determined by the surface glycoprotein hemagglutinin (HA) and amino acid substitutions at exposed epitope sites in HA mediate loss of recognition by antibodies. Here, we show that antigenic differences measured through serological assay data are well described by a sum of antigenic changes along the path connecting viruses in a phylogenetic tree. This mapping onto the tree allows prediction of antigenicity from HA sequence data alone. The mapping can further be used to make predictions about the makeup of the future seasonal influenza virus population, and we compare predictions between models with serological and sequence data. To make timely model output readily available, we developed a web browser based application that visualizes antigenic data on a continuously updated phylogeny.

Seasonal influenza viruses evade immunity in the human population through frequent amino acid substitutions in their hemagglutinin (HA) and neuraminidase (NA) surface glycoproteins (Hay *et al.*, 2001). To maintain efficacy, vaccines against seasonal influenza viruses need to be updated frequently to match the antigenic properties of the circulating viruses. To facilitate informed vaccine strain selection, the genotypes and antigenic properties of circulating viruses are continuously monitored by the World Health Organization (WHO) Global Influenza Surveillance and Response System (GISRS), with a substantial portion of the virological characterizations being performed by the WHO influenza Collaborating Centers (WHO CCs) (Barr *et al.*, 2014).

Antigenic properties of influenza viruses are measured in hemagglutination inhibition (HI) assays (Hirst, 1942), that record the minimal antiserum concentration (titer) necessary to prevent crosslinking of red blood cells by a standardized amount of virus based on hemagglutinating units (HAU). An antiserum is typically obtained from a single ferret infected with a particular reference virus. For a panel of test viruses, the HI titer is determined by a series of two-fold dilutions of each antiserum. An antiserum is typically potent against the homologous virus (the reference virus used to produce the antiserum), but higher concentrations (and hence lower titers) are frequently required to prevent hemagglutination by other (heterologous) viruses. HI titers typically decrease with increasing genetic distance from the homologous virus ((Hay *et al.*, 2001)).

Given multiple antisera raised against different reference viruses and a panel of test viruses, WHO CCs routinely measure the matrix of distances between a set of antisera and multiple test viruses, see Fig. 1A. This distance matrix can be visualized in two dimensions via multidimensional scaling – an approach termed antigenic cartography (Smith *et al.*, 2004). While standard cartography does not use sequence information, sequences have been used as priors for distances in a Bayesian version of multidimensional scaling (Bedford *et al.*,
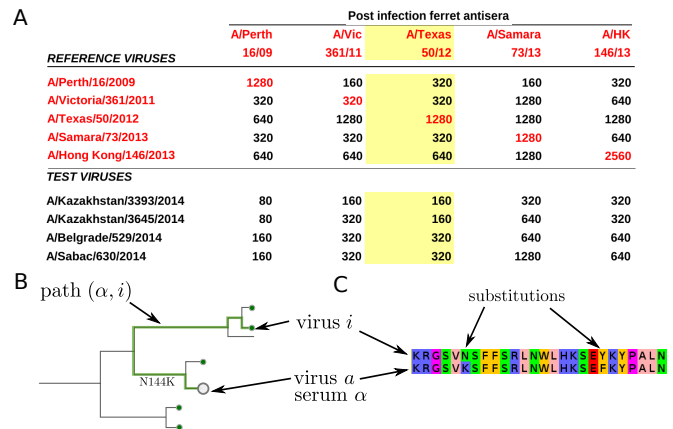

| | Post infection ferret antisera | | | | |
REFERENCE VIRUSES	A/Perth 16/09	A/Vic 361/11	A/Texas 50/12	A/Samara 73/13	A/HK 146/13
A/Perth/16/2009	1280	160	320	160	320
A/Victoria/361/2011	320	320	320	1280	640
A/Texas/50/2012	640	1280	1280	1280	1280
A/Samara/73/2013	320	320	320	1280	640
A/Hong Kong/146/2013	640	640	640	1280	2560
TEST VIRUSES					
A/Kazakhstan/3393/2014	80	160	160	320	320
A/Kazakhstan/3645/2014	80	320	160	640	320
A/Belgrade/529/2014	160	320	320	640	640
A/Sabac/630/2014	160	320	320	1280	640

FIG. 1 **Antigenic data and models for HI titers**. A: A typical table reporting HI titer data. Each number in the table is the maximum dilution at which the antiserum (column) inhibited hemagglutination of red blood cells by a virus (row). The red numbers on the diagonal indicate homologous titers. A typical HI assay consists of all reciprocal measurements of the available antisera and reference viruses, and a number of test viruses that are measured against all antisera, but for which no homologous antiserum exists. B: Each HI titer between antiserum α and virus i can be associated with a path on the tree, indicated as a thick line. The tree model seeks to explain the titer distance as additive contributions of branches. C: In the substitution model, the sum over branches on the tree is replaced by a sum of contributions of amino acid substitutions.

2014). In order to infer contributions of individual amino acid substitutions to antigenic evolution, Harvey *et al.* (2014); Sun *et al.* (2013) have used models that predict HI titer differences by comparing sequences of reference and test viruses.

Here, we show that antigenic distance relationships are accurately described by a model based on phylogenetic tree structure. We use the model to show that HI distances have a largely symmetric and tree-like structure. We show that large effect mutations account for about half of the total antigenic change and the effect of specific substitutions is strongly dependent on the genetic background in which they occur. We further investigate the ability of HI measurements to predict dominant clades in the next influenza season. To visualize antigenic properties on the phylogenetic tree, we have integrated the models of antigenic distances and the raw HI titer data into nextflu – an interactive real-time tracking tool for influenza virus evolution (Neher and Bedford, 2015).

This comprehensive summary of HA sequences from past and current influenza viruses linked to their antigenic properties has the potential to inform vaccine strain selections and facilitate efforts to predict successful influenza lineages (He and Deem, 2010; Łuksza and Lässig, 2014; Neher *et al.*, 2014; Steinbrück *et al.*, 2014; Steinbrück and McHardy, 2011).

Results

We use two related models that predict HI titers from sequences. The first – the *tree model* – expresses logarithmic HI titers by a sum of titer drops along branches in the phylogenetic tree that connect the test virus and the reference virus against which the antiserum was raised, see Fig. 1B. The second – *the substitution model* – explains HI titers as a sum of titer drops associated with amino acid substitutions between reference and test virus, see Fig. 1C. Not all branches and not all substitutions are expected to result in changes to titer. These two models are similar but complement each other in a few aspects that we discuss further below. In addition to titer drops associated with branches or substitutions, individual antisera and viruses differ in their activity in HI assays. In the model, we account for this variability through 'antiserum potency', which raises or lowers the expected titer of all HI measurements against an antiserum, and 'virus avidity', which raises or lowers the expected titer of all HI measurements against a virus.

Specifically, in the tree model the antigenic distance $\Delta_{i\alpha}$ between virus i and antiserum α is modelled as

$$\Delta_{i\alpha} = p_\alpha + c_i + \sum_{b\in(i...\alpha)} d_b \qquad (1)$$

where d_b, p_α, c_i denote the titer drop on a branch b in the path $(i \ldots \alpha)$, the potency of antiserum α, and the avidity of virus i, respectively. In the substitution model, the sum over branches is replaced by a sum over amino acid substitutions between reference and test virus. The parameters of our model are determined by minimizing a cost function that encourages a sparse model, i.e., a model that explains titers with a small number of large parameters and many parameters set to zero.
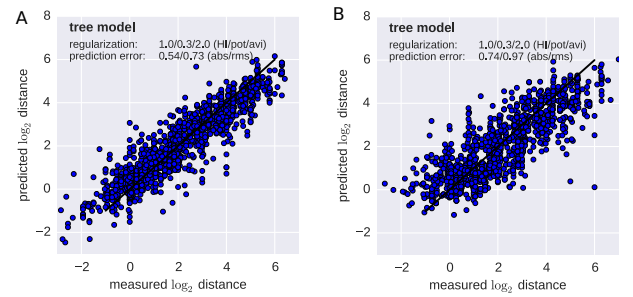

FIG. 2 **HI titers are accurately predicted by the tree model**. The figure scatters predicted (y-axis) against a test set of measurements not used for training of the model (x-axis). This test set either consists of A) a random sample of 10% of all measurements, or B) all measurements for 10% of all viruses. In the latter, no avidity of a virus can be estimated for viruses in the test set since these viruses are completely absent from the training data. Hence prediction accuracy is lower but still comparable to the measurement accuracy.

TABLE I Prediction accuracy of the tree and substitution model. The terms *titer* and *virus* refer to predictions of individual titer measurements or all measurements for viruses completely absent from the training data, respectively.

model:	tree		substitution	
test set:	titer	virus	titer	virus
A(H3N2) (12y)	0.54	0.72	0.53	0.71
A(H1N1pdm09) (7y)	0.46	0.77	0.5	0.72
B/Yam (12y)	0.68	0.86	0.65	0.88
B/Vic (12y)	0.74	0.86	0.76	0.84

This is achieved by ℓ_1-norm regularization and non-negativity constraints, corresponding to the assumption that antigenic effects are positive and exponentially distributed. In contrast, the parameters absorbing variation among virus/antisera interactions are regularized by their ℓ_2-norm, corresponding to a Gaussian prior on these parameters. For a detailed description of the models and inference procedures, see Material and Methods.

The decomposition of the HI data into antiserum potency, avidity of a virus, and a sparse genetic component (branches or substitutions) has the effect of smoothing the HI distance relationships and virus and antiserum specific variation is absorbed into the corresponding effects. The substitution or branch effects pick up titer drops associated with a larger number of antiserum-virus pairs.

The tree and substitution models accurately predict HI titers

To evaluate the performance of the models we trained them on 90% of the data and used the remaining measurements to validate the models as in (Bedford *et al.*, 2014). We found that the models were able to predict titers of antiserum-virus com-
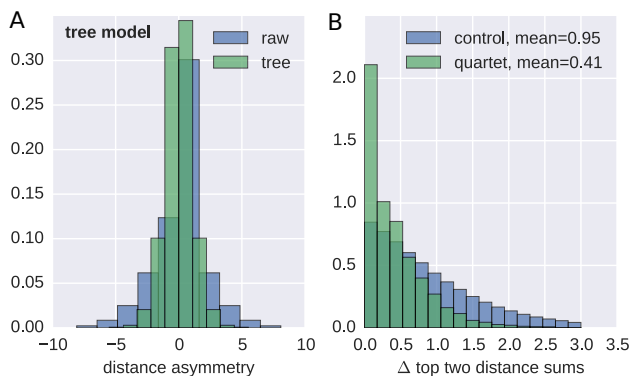
FIG. 3 **HI titers approximate a distance on the tree.** A) Reciprocal HI titer measurements often differ by several $\log_2$ units. After subtracting the inferred virus and antiserum effects the remaining HI titer distance is almost symmetric with deviations again on the order of the measurement accuracy. B) On a tree, the two largest sums of distances within antisera / virus quartets are equal (see text). The figure shows the distribution of the absolute value of the difference of the top two distance sums for quartets and for three random distance sums. Those from quartets have a much smaller deviation.

binations to an accuracy of approximately $0.5 \log_2$ titer levels for H3N2 with somewhat lower accuracy for the influenza B lineages (Tab. I, Fig. 2A). Very little consistent antigenic evolution is observed in A(H1N1pdm09).

To quantify the prediction accuracy for viruses for which no antigenic data exists, we selected 10% of the viruses and excluded all measurements involving these viruses from the training data. Both models predicted titers for viruses not part of the training set to an accuracy of approximately $0.75 \log_2$ titer levels (Tab. I, Fig. 2B). Having completely excluded a virus from the training data implies that no avidity of a virus can be estimated. The increased prediction error is therefore largely due to virus-to-virus variability that is not captured by the HA phylogeny.

In order to infer the genetic component of an HI titer, the relevant branches in the tree or the substitutions that separate test and reference virus have to be constrained by measurements in the training data set. For a completely novel clade in the tree, the model would predict HI titers equal to that of the base of the clade for all subtending viruses. Similarly, accurate inferences by the substitution model require the effects of the relevant substitutions to be constrained by training data.

Using the tree and substitution models, we can predict HI titers for every combination of antiserum and virus in a phylogenetic tree (with prediction confidence varying by quality and amount of antigenic data), essentially providing a smoothed interpolation of antigenic evolution on the tree. Note that the model correctly predicts titers in excess of homologous titers (negative values in Fig. 2A). These higher titers often coincide with large negative virus avidities, which explains the absence of titers predicted to be strongly negative in Fig. 2B, where no virus avidities are available.

HI titers approximate a symmetric tree distance

To further evaluate the approach of partitioning a titer into a antiserum, a virus, and a tree component, we considered all reciprocal titer measurements, i.e., pairs of viruses a, b against which antisera α, β have been raised and that have been measured against each other. Subtracting the virus avidities and antiserum potency contributions from titers, the remainders $H_{a,\beta} - c_a - p_\beta$ and $H_{b,\alpha} - c_b - p_\alpha$ should reduce the titers to the symmetric tree component. Fig. 3A compares the distribution of $H_{a,\beta} - c_a - p_\beta - H_{b,\alpha} + c_b + p_\alpha$ with the uncorrected difference between the reciprocal titers $H_{a,\beta} - H_{b,\alpha}$. While raw reciprocal titer measurements often differ by several $\log_2$ titer units, the corrected tree component was symmetric to within one unit, as expected.

The degree to which titer distances have tree-like properties can be tested using the following quartet rule: take any four leaves a, b, c, d and construct the sum of the distances $d_{ab} + d_{cd}$, $d_{ac} + d_{bd}$ and $d_{ad} + d_{bc}$. If the distances are given by a tree, the two largest of the three sums of distances will be equal. To test this quartet rule we determined maximal 'cliques' of reference antisera, the activities of which had been measured against all viruses used to generate the antisera. Out of these cliques, we selected at random four antisera, and compared the two largest distance sums. As a control, we constructed three sums of two distances by three times randomly drawing four antisera. The difference between the largest and second largest of these distance sums has a mean of $0.41 \log_2$ titer levels, while the control has a mean of $0.95 \log_2$ titer levels (Fig. 3B). The similarity of the largest and second largest sums supports a tree-like structure underlying observed HI distances.

Amino acid substitutions associated with titer drops

The majority of antigenic evolution tends to occur at a subset of sites (Muoz and Deem, 2005; Shih *et al.*, 2007) with seven positions near the HA receptor binding site (Koel 7) playing a particularly prominent role (Koel *et al.*, 2013). However, the number of substitutions at these sites is a poor predictor of antigenic distance with r^2 values of about 0.25 (supplementary Fig. S1), possibly because the effect of a substitutions depends on the genetic background (HA sequence) in which it occurs.

Within the substitution model, we can explicitly associate antigenic changes with amino acid substitutions. Tab. S1 shows the top 50 contributions for A(H3N2) for sequences estimated from sets of sequences and the associated HI titers in five overlapping 10 year intervals. Most of the largest contributions coincide with substitutions at Koel 7 sites. When a substitution is present in the same context in overlapping time intervals, the estimated effects tend to be similar (K189N, K135E, K158R, K140E, Y159F) – provided there is enough data to constrain the model.

By and large, the effects we infer are compatible with those associated with cluster transitions determined in (Koel *et al.*, 2013). For example, the transition from the Sichuan/87 (SI87)
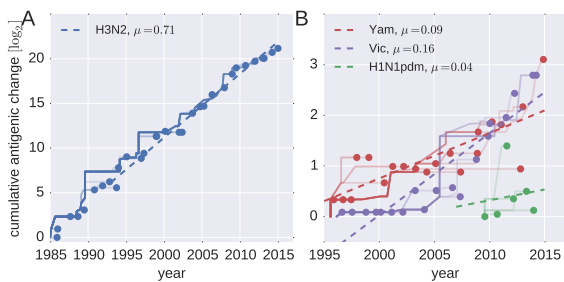
FIG. 4 **Cumulative antigenic change from the root to the leaves of the tree.** A) shows H3N2 antigenic evolution over the past 30 years. B) shows the corresponding traces for the two influenza B lineages (past 20 years) and H1N1pdm09 (past 7 years).

cluster to the Beijing/89 (BE89) cluster involved the substitution N145K, for which we infer an effect of 1.95 antigenic units (each unit is equal to a two-fold decrease in HI titer). Together with minor effects associated with N193S, I186S, and G135N, this comes close to the map distance 3.9 units. For the transition from SI87 to Beijing/92 (BE92), the substitution model estimates a distance of 4.5 that is associated with many almost simultaneous substitutions – hence the exact assignment remains difficult to pin down. The map distance corresponding to this cluster transition is 7.8 units, while the typical titer drop is approximately 5 units and hence closer to the estimate of the tree or substitution model. The transition from BE92 to Wuhan/95 (WU95) is gradual with several substitutions of intermediate magnitude. From WU95 to Sydney/97 (SY97), we estimate a distance of approximately 3 units of which 2 units are attributed to the set of substitutions: K62E/V144I/K156Q/E158K/V196A/N276K. These substitutions account for 3.6 units in cartography. Lastly, the transition from SY97 to Fujian/02 (FU02) is attributed to Q156H (1.05 units as in (Koel *et al.*, 2013)) and N121T. Most of the distances we infer between clusters are smaller than those estimated in cartography, which is compatible with our finding of overall reduced antigenic drift when modeling antigenic evolution in many dimensions (see below).

The inferred effects of substitutions suggest substantial dependence on genetic background and the specific amino acid change. For HA1 position 145, the back and forth between asparagine (N) and lysine (K) is associated with different effects in 1989–1992, and 2004. Interestingly, the inferred effect of the forward substitutions (N145K in 89–92 and K145N in 04) is large in these instances, while the inferred effect for the backward direction is small. Some substitutions at Koel 7 positions have very small effects, as for example E158D and S189N which involve biochemically similar amino acids. Other substitutions at these positions have large effects, but don't spread. K158R, for example, shows up repeatedly and is associated with a 2 unit titer drop without ever reaching high frequencies. A full table of all inferred effects for substitutions at particular positions in overlapping 10 year intervals is given as supplementary Tab. 2. However, the interpretation of effects is sometimes difficult due to colinearities between substitutions.

Cumulative antigenic evolution

By summing all contributions to antigenic change on branches on the path between a virus and the root of the tree, we can estimate the total past antigenic change. This cumulative antigenic change is roughly comparable to 'dimension 1' in antigenic cartography (more precisely analogous to the length of a path on the map corresponding to the trunk of the tree). H3N2 viruses advance by approximately 0.7 $\log_2$ titer units per year (Fig. 4A), while influenza B virus lineages advance 0.16 (Vic) and 0.09 (Yam) units per year (Fig. 4B). Although consistent in relative rates across lineages, these absolute rates tend to be slightly lower than estimated previously using cartography (Bedford *et al.*, 2014), possibly because forcing the antigenic distance matrix into two dimension causes distortions resulting in a slight exaggeration of the distances. Very little consistent antigenic evolution is observed in H1N1pdm09 (Fig. 4B).

We estimate that in A(H3N2) approximately half the cumulative antigenic evolution is due to a large number of substitutions with effects smaller than one unit, while 20% is accounted for by a few substitutions with effects above two units (Fig. S3). However, some of largest effects are associated with clusters of colinear amino acid changes and their individual effects cannot be resolved.

Limitations of the models

The tree and substitution model perform similarly in terms of accuracy as summarized in Tab. I, which is expected since branches of the tree are associated with substitutions and vice versa. In particular circumstances, however, one model is more accurate than the other and the two models complement each other in a number of situations.

The tree model assumes that titers are additive along the tree. While this is in general a reasonable assumption, it is violated when the same amino acid position is mutated multiple times in different parts of the tree. In a recent example from H3N2, there are independent substitutions at HA1 site 159 with F159Y resulting in clade 3c2.a viruses and F159S resulting in clade 3c3.a viruses. The distance between 3c2.a and 3c3.a viruses is not necessarily a sum of the effects associated with the branches on which these two substitutions occur. In such cases, the substitution model tends to be more accurate as it has the additional freedom to introduce the Y159S substitution (and the reverse) that directly compares 3c2.a and 3c3.a viruses.

Similarly, the substitution model fails when the same substitution (as opposed to different substitutions at the same site) occurred in different places on the tree in different genetic backgrounds. For sequence ensembles covering long time periods, a substitution might have happened at different times in different genetic backgrounds. The model will fit a single effect, even though the effect of the substitution might be background dependent. Furthermore, the substitution model tends to be inaccurate when predicting titers for test viruses that predate the reference virus. Such 'back-in-time' measure-
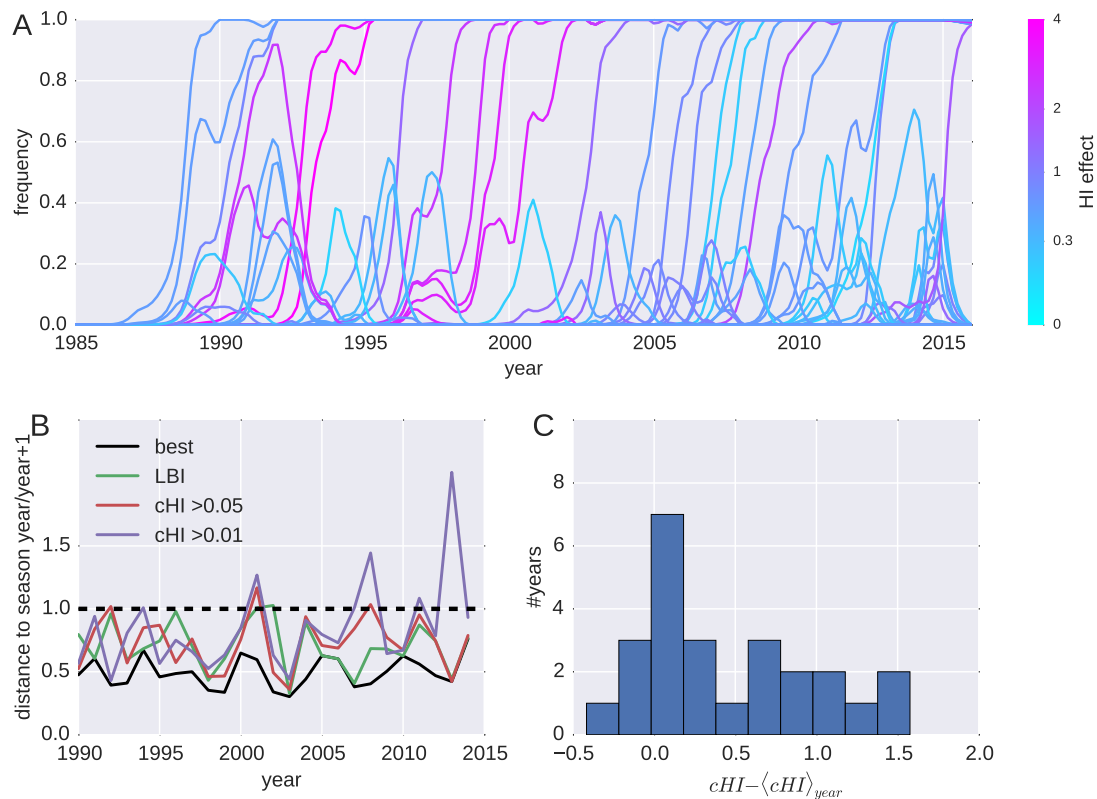
FIG. 5 **Antigenic evolution and the success of clades.** A) shows trajectories of clades colored by inferred cumulative antigenic evolution in the past 6 months on branches ancestral to the clade. The majority of the clades with very large HI effects fix, but several branches with substantial antigenic evolution go extinct. B) shows genetic distances to the next season of nodes with maximal LBI or cHI (distance is normalized such that the average clade has distance 1). The black line indicates the distances of the best possible pick. Predictions based on maximal LBI and cHI are correlated, but maximal cHI is sensitive to false positives. C) shows the histogram of centered cHI of the clade closest to the next season for each year, suggesting that successful clades tend to be antigenically advanced.

ments are underrepresented in the data, and while the forward substitution might have a large effect assigned, the few back-in-time measurements do not provide enough support to include the reverse substitutions into the model. The tree model does not suffer from this problem, as effects are assumed to be symmetric.

By and large the model accurately predicts measured HI titers (Fig. 2) and deviations affect only isolated clades, typically when very few measurements are available to constrain the model. The visualization described below allows a direct side by side comparison of the two models and the measurements, which makes it easy to identify such isolated inaccuracies.

Antigenic change and the success of clades

Antigenic changes result in viruses able to reinfect individuals with immunity to previously circulating viruses. Intuitively, large antigenic changes should therefore be positively selected for and rapidly spread through the virus population, unless there is a fitness cost for the virus. We investigated the relationship between the amount of antigenic change and suc-

cess of clades in the phylogenetic tree. Fig. 5 shows frequency trajectories of clades reaching at least 10% at one time as estimated by nextflu, while color indicates the magnitude of the antigenic change that accumulated along the ancestral lineage over the past six months. Although large antigenic changes appear to fix more often than small antigenic changes, there are also several clades that evolved antigenically but failed to spread.

For each season, we determined the clade with highest Local Branching Index (LBI, a predictor of clade success (Neher *et al.*, 2014)) and the clade with the largest antigenic advancement (cHI) relative to all other viruses in a season (restricted to clades that account for at least 5% of available sequences for the given season). For each of these clades, Fig. 5B shows genetic distance to the virus population of the following season. Antigenic advance as measured by cHI is predictive of which lineage would dominate the following season: the distance is significantly below the population average. Predictions by cHI and LBI are comparable in quality and are correlated. Yet, clades with maximal cHI are sometimes far from the future population suggesting that a predictor based on antigenic phenotype alone readily generates false positives: the problem becomes worse when smaller clades are included (threshold
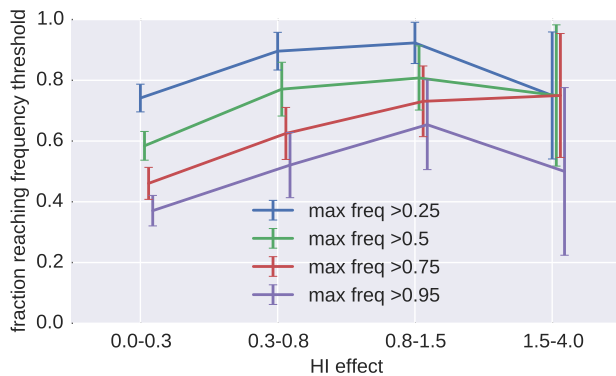
FIG. 6 **Substitutions with large antigenic effect fix preferentially**. The figure shows the fraction of all substitutions that reach the indicated frequency for different magnitudes of the inferred antigenic effect of this substitution. The plot combines data from substitution models fitted to 5 overlapping 10 year intervals from 1985 to 2015 and contains all substitutions that reach at least 10% population frequency. There are few substitutions in the highest HI category – error bars show the standard deviation over boot strap replicates of substitutions.

FIG. 7 **Visualization of antigenic evolution on the tree**. Tree tips are colored by the predicted $\log_2$ titer distance relative the focal virus (A/Texas/50/2012 in this case). The tool tip shows titers relative to all available antisera and titer predictions as a table. Each reference virus (grey squares) can be chosen by clicking with the mouse, upon which the tree color is updated. Crosses indicate vaccine strains. The tool tip shows all available measurements and the predictions by both models, while the color averages measurements relative to different antisera.

lowered to $> 1\%$ of all sequences). Nevertheless, successful clades tend to be antigenically advanced. Fig. 5C shows the histogram of centered cHI for clades closest to the next season. These clades tend to have larger cHI than co-circulating viruses. These results suggest that cHI correlates with viral fitness but is not the only factor determining clade success. Supplementary Fig. S2 compares the cHI of the most antigenically advanced clade to that of the successful clade in each year.

As an alternative to clades in the tree, Fig. 6 shows the probability of a substitution reaching a maximal frequency as a function of its inferred antigenic effect. This fixation probability increases with antigenic effect, but even substitutions with very large effect can fail to spread – an effect which limits its predictive power.

Visualization of antigenic evolution

Antigenic evolution can be visualized by mapping titer distances into the plane using a variant of multi-dimensional scaling (Smith *et al.*, 2004). This two dimensional representation, however, is not readily superimposed with the sequence evolution of the viruses. Instead of squeezing the sequence evolution onto the plane (Bedford *et al.*, 2014), we map the HI titer data onto the phylogenetic tree using nextflu (Neher and Bedford, 2015).

The application nextflu tracks in near real time the evolution of seasonal influenza viruses and allows users to explore recent changes at particular positions, spot rapidly growing clades, and analyze the geographic distribution of viruses. We have integrated the tree and sequence models of titer data into nextflu's Python based processing pipeline augur. The titer data and the models are exported along with the tree and visualized using auspice, the JavaScript based front-
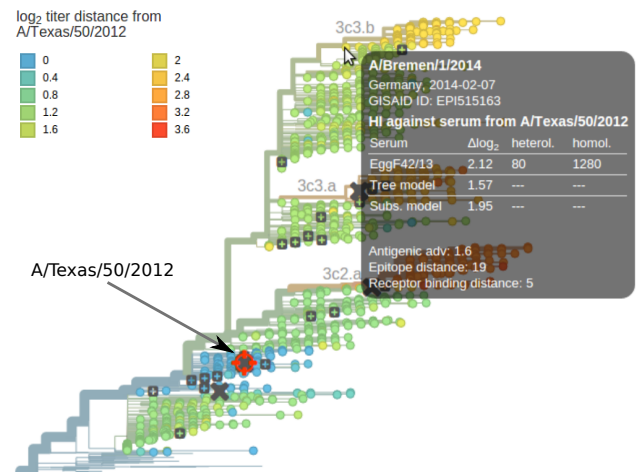
end of nextflu. The resulting visualization is available at `HI.nextflu.org` and Fig. 7 shows a screen shot.

This web visualization allows exploration of HI distance relative to specific antisera. All reference viruses against which antisera have been raised are indicated by grey squares. A *focal reference virus* can be chosen by clicking on one of these squares, thereby coloring the tree by the average antigenic distance between viruses and antisera raised against the focal virus. Tooltips – information boxes that pop up when the mouse hovers over a virus – display all available measurements along with the predicted titers relative to the focal virus. The normalized and log scaled titers can be optionally corrected using estimates of antiserum potency and virus avidity.

The tree can be colored either by measured titers (which are available only for a subset of the viruses) or the predictions by either the tree model or the substitution model. Toggling between coloring by measured and predicted titers gives an intuitive visual impression of the noise in titer data and possible inaccuracies of the model predictions. The model predictions are essentially de-noised antigenic distances that integrate information from a large number of titer measurements.

In addition to titer measurements and titer predictions, the `HI.nextflu.org` website allows coloring of the tree by the antigenic evolution that accumulated along branches starting from the root of the tree. The latter is similar to dimension one in antigenic cartography.

Discussion

We have shown that antigenic evolution of seasonal influenza viruses can be accurately predicted from HA sequences using models parameterized by branches separating test and reference virus in a phylogenetic tree or by amino acid substitutions separating the two sequences. Both of these models predict titers with similar or better accuracy than more traditional cartographic approaches. The appeal of antigenic cartography (Smith *et al.*, 2004) is partly the ease with which complex tables of HI titer data can summarized, visualized, and printed on paper. To achieve similarly intuitive display of the results of the tree and substitution models, we implemented a web-browser based interactive visualization available at `HI.nextflu.org` (Neher and Bedford, 2015). The web application integrates antigenic relationships with molecular evolution, geographic distribution, and predictors of clade growth. Our goal, rather than simply predicting antigenic data from genetic sequences, was a characterization of antigenic space, a full integration with the phylogenetic tree, and an assessment of the predictive value of HI measurements. In contrast, Harvey *et al.* (2014) focus on identifying specific residues in H1N1 that are responsible for titer drops, while (Sun *et al.*, 2013) focus on predicting titers of unknown viruses visualized via cartography.

Previous analysis had concluded that two dimensions provide the optimal embedding for antigenic evolution and adding additional dimensions did not improve the predictive power (Bedford *et al.*, 2014; Smith *et al.*, 2004) Here we find that a model based on the tree structure – effectively infinite dimensional – predicts titers similar or better accuracy. This apparent discrepancy has its roots in the number of parameters necessary to specify the model. In d dimensional cartography, the location of each virus is described by d coordinates. Hence the number of parameters increases rapidly with the number of dimensions of the model and predictive power decreases due to overfitting. However, if distances are additive and tree-like, all $n(n-1)/2$ pairwise distances between n viruses are defined by the contributions of at most $n-2$ internal and n external branches (external branches are not considered by the model). In practice, the number of internal branches is substantially lower than $n-2$ due to the many polytomies in the tree. For A(H3N2) from the past 12 years, HI data was available for 1796 out of total of 2473 viruses in the tree. Of the 622 internal branches constrained by HI data, 189 branches were inferred to affect HI titers (¿0.001). Hence with substantially fewer parameters, we achieve a better or comparable fit to the data than $d = 2$ cartography, which suggests that the phylogenetic tree is a more natural space for antigenic change. We corroborated this interpretation by explicitly testing "tree-ness" using quartet distances and symmetry between reciprocal measurements (Fig. 3).

We find substantial differences in the overall rate of antigenic evolution across viruses with fast antigenic drift in H3N2 and slow antigenic drift within both influenza B lineages, in agreement with Bedford *et al.* (2014). Overall, antigenic drift estimated by our method tends to be somewhat lower, in particular for influenza B. This discrepancy in drift is likely due to the different model space: the phylogenetic tree provides more freedom for different clades to evolve in different directions and antigenic distances are accommodated on side branches, rather than the trunk. On a two dimensional antigenic map, however, the space for side branches and subclusters is limited, such that more antigenic distance is picked up by the backbone of the map.

Current efforts to predict the evolution and dynamics of seasonal influenza viruses (He and Deem, 2010; Łuksza and Lässig, 2014; Neher *et al.*, 2014; Steinbrück *et al.*, 2014; Steinbrück and McHardy, 2011) are based solely on virus HA sequences. By mapping the phenotypic HI data onto sequences and phylogenetic trees, it should be possible to improve prediction accuracy. Using HI data to predict is, however, not as straightforward as it might seem and by itself it does not predict better than the sequence-based LBI predictor (Neher *et al.*, 2014). HA substitutions associated with large antigenic changes have a higher probability of fixation, but many causing substantial antigenic change (e.g. K158R) fail to spread in the population. Similarly, in many years clades that are antigenically more distant from previously circulating viruses die out. The high frequency of such "false positives" interferes with the use of HI measurements for early detection of emerging strains that are the likely founders of future generations of the virus. We observe substantial false positives at the 1% clade frequency level, but fewer false positives at the 5% frequency level (Fig. 5B). However, using a 5% frequency threshold loses much of the early detection capacity, limiting HI-based prediction to the regime accessible to the sequence-based genealogy approaches, such as LBI.

The main challenge is hence to find a way to reduce the false positive rate in HI-based prediction. Success of strains with smaller HA antigenic advancement over the ones with a larger advancement could be rationalized in two ways: i) by some alternative improvement of infectivity, immune system avoidance or lower mutational load; ii) by a fitness cost associated with a large antigenic effect substitutions. Supporting the second scenario, it is known that adaptive mutations are sometimes not tolerated in certain genetic backgrounds because they destabilize the encoded protein and further, stabilizing, mutations are required to compensate for the loss of virus fitness (Gong *et al.*, 2013; Thyagarajan and Bloom, 2014). Better understanding of the context dependence of large antigenic effect HA substitutions may therefore be a promising way towards reducing the false positive rate and improving prediction capacity. The problem of false positive detection is also seen at the Koel 7 positions. While most past dramatic changes in antigenic phenotype are associated with substitutions at these sites, substitutions at Koel 7 sites do not always have antigenic effects or fail to spread. Koel 7 substitutions alone are poor predictors.

In conclusion: our study demonstrates that HI data integrates naturally onto the sequence derived phylogeny of the virus. While at present HI-based prediction does not outperform sequence based methods, better understanding of genetic context dependence of HI data may provide a way towards improved performance. Characterizing HA substitutions that have historically been associated with antigenic transitions

and placing HI data directly into genealogical context may also help with optimizing targeted acquisition of HI data.

Materials and Methods

Data

HA sequences of influenza A and B viruses isolated from humans were downloaded from GISAID. Accession numbers of all sequences are provided as supplementary files. We collected HI data from publications Barr *et al.* (2010); Russell *et al.* (2008); Smith *et al.* (2004) and annual and interim reports of the WHO CC London between 2002 and 2015 (Hay *et al.*, 2008, 2009, 2002; McCauley *et al.*, 2011, 2010a,b, 2012a, 2013a,b, 2014a,b, 2015, 2012b). HI data prior to 2011 was curated in Bedford *et al.* (2014)). Original HI data tables are available from the website of the Worldwide Influenza Centre at the Francis Crick Institute at http://www.crick.ac. uk/research/worldwide-influenza-centre/ annual-and-interim-reports/.

While the modalities of HI assays have changed over the years (red blood cells from different species, addition of neuraminidase inhibitors, etc), we find that the model describes data spanning many years with reasonable accuracy. This insensitivity of the model is likely due to the fact that differences in HI assay methodologies can largely be absorbed in the model terms for antiserum potency and virus avidity. However the most recent data, largely provided by WHO CC London, is modeled with greater accuracy.

Data processing and model fit

The data processing pipeline is based on **nextflu** (Neher and Bedford, 2015), which subsamples viruses, aligns sequences and builds a phylogenetic tree. This pipeline was modified for the current purpose to enforce the inclusion of all strains for which antigenic data was available. Then, in addition to the standard **augur** pipeline of nextflu, the tree and substitution models were fitted to the HI data as follows.

For each combination of virus i and antiserum α, we define antigenic distance as $H_{i,\alpha} = \log_2(T_{a\alpha}) - \log_2(T_{i\alpha})$, where $T_{i\alpha}$ is the antiserum titer required to inhibit virus i and $T_{a\alpha}$ is the homologous titer. In case multiple measurements are available, we average the base 2 logarithm of the titers. When no homologous titer was available, the maximal titer was used as a proxy for the homologous titer. The path between virus i and antiserum α extends over branches of the phylogeny $b \in (i \ldots \alpha)$, where each branch has a branch effect d_b. The tree distance $\Delta_{i,\alpha}$ between virus i and antiserum α is defined in Eq. (1). The parameters d_b, p_α, c_i are then estimated by minimizing the cost function

$$C = \sum_{i,\alpha}(H_{i,\alpha} - \Delta_{i,\alpha})^2 + \lambda \sum_b d_b + \gamma \sum_i c_i^2 + \delta \sum_\alpha p_\alpha^2 \quad (2)$$

subject to the constraints $d_b \geq 0$. To avoid overfitting, the different parameters of the model are regularized by the last three terms in Eq. (2). Large titer drops are penalized with their absolute value multiplied by λ (ℓ_1 regularization), which results in a sparse model in which most branches have no titer drop (Candes and Tao, 2005). Similarly, the antiserum and virus avidities are ℓ_2-regularized by γ and δ, penalizing very large values without enforcing sparsity. This constrained minimization can be cast into a canonical convex optimization problem and solved efficiently, see below. In the substitution model, the sum over the path in the tree is replaced by a sum over amino acid differences in HA1. Sets of substitutions that always occur together are merged and treated as one compound substitution. The inference of the substitution model parameters is done in the same way as for the tree model (see Harvey *et al.* (2014); Sun *et al.* (2013) for a similar approach). Since there are only a small number of antisera and differences in antiserum potency are often on the order of one or two antigenic units, δ was assigned a small value of 0.3, while $\lambda = 1.0$ was used to regularize branch or substitution and $\gamma = 2.0$ for virus effects. The quality of the fit depends weakly on these parameters.

The total number of adjustable parameters are S antiserum potencies, V avidities of viruses, and M internal branches of the tree of which there are at most $V - 1$, but typically fewer due to many polytomies in the trees. In practice only a fraction of the branches have non-zero branch effects and the total number of non-zero parameters is not much larger than the number of test viruses. In the substitution model, the number of non-singleton substitutions found in an HA1 alignment is typically on the order of 100, most of which are inferred to have no antigenic effect.

This optimization of Eq. (2) can be cast into a canonical quadratic programming problem of the form

$$\begin{aligned} \text{minimize}: & \quad \tfrac{1}{2}\mathbf{x}\mathbf{Q}\mathbf{x} + \mathbf{q}\mathbf{x} \\ \text{subject to}: & \quad \mathbf{G}\mathbf{x} \leq \mathbf{h} \end{aligned} \quad (3)$$

where $\mathbf{x}$ is the vector of unknowns, and the matrix $\mathbf{Q}$ and the vector $\mathbf{q}$ specify the cost function. The matrix $\mathbf{G}$ and the vector $\mathbf{h}$ encode inequality constraints on $\mathbf{x}$.

To formulate Eq. (2) in this canonical form, we concatenate the titer drops d_b associated with branches b of the tree, the potencies p_α of each antiserum, and avidities c_i of virus isolates into a single vector $\mathbf{x}$

$$x_s = \begin{cases} d_s, & \text{for } s = 1 \ldots B \\ c_s, & \text{for } s = B+1 \ldots B+V \\ p_s, & \text{for } s = B+V+1 \ldots B+V+S \end{cases} \quad (4)$$

where B are the number of branches in the tree connecting measurements, V is the number of viruses with titer measurements, and S are the number of antisera. Next, we construct a large binary matrix $\mathbf{A}$ of dimension $N \times (B+V+S)$, where N is the total number of measurements. Each row of the matrix $\mathbf{A}$ codes for a titer prediction $\Delta_{i\alpha}$. Using the double index $i\alpha$ to label measurements, entries of $\mathbf{A}$ are given by

$$\mathbf{A}_{i\alpha,s} = \begin{cases} 1, & \text{for } s \text{ in path, potency, or avidity} \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

i.e., $\mathbf{A}_{i\alpha,s} = 1$ for all s that correspond a branch of the path $(i \ldots \alpha)$, the virus avidity c_i and the antiserum potency p_α; otherwise, $\mathbf{A}_{i\alpha,s} = 0$. All titer predictions of Eq. (1) are hence given by

$$\mathbf{\Delta} = \mathbf{A}\mathbf{x} \tag{6}$$

where $\mathbf{x}$ is the vector of parameters.

Using these definitions, the cost function Eq. (2) can be written as

$$(\mathbf{H} - \mathbf{A}\mathbf{x})^T(\mathbf{H} - \mathbf{A}\mathbf{x}) + 2\lambda \sum_{s=1}^{B} x_s + \gamma \sum_{s=B+1}^{B+V} x_s^2 + \delta \sum_{s=B+V+1}^{B+V+S} x_s^2 \tag{7}$$

subject to $x_s \geq 0$ for $s = 1 \ldots B$. Dropping constant terms and defining

$$Q_{rs} = \begin{cases} \sum_i A_{ir} A_i s, & \text{if } r, s \leq B \\ \gamma & \text{if } r = s, s > B, s \leq B + V \\ \delta & \text{if } r = s, s > B + V \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

we have

$$\mathbf{x}\mathbf{Q}\mathbf{x} - 2\mathbf{H}\mathbf{A}\mathbf{x} + 2\lambda \sum_{s=1}^{B} x_s \tag{9}$$

To enfore the ℓ_1 regularization and the positivity of the titer drops corresponding to x_s, $s = 1 \ldots B$, we define inequality constraints

$$x_s \geq 0 \tag{10}$$

for $s = 1 \ldots B$ which forces all titer drops to be positive. In addition, we set

$$\mathbf{q} = -\mathbf{H}\mathbf{A}\mathbf{x} \tag{11}$$

and add λ to q_s, $s = 1, \ldots B$ to penalize large effects. With these definitions, we have cast Eq. (2) in the form of Eq. (3). The resulting quadratic programming problem is then solved with cvxopt by M. Andersen and L. Vandenberghe.

The HI titer data and the inferred model parameters are integrated into the json data structure describing the tree or saved in an additional data file for later visualization using **auspice**.

Visualization

The HI titer coloring and tool tips is implemented via straightforward extension of nextflu/auspice functionality. In addition to the standard nextflu tree display, a structure showing the positions at which the substitution model inferred large contribution to antigenic change are shown on the pages for each individual virus lineage. The structures are visualized with JSmol (JSmol developers, 2015). For H3N2, we use structure 5HMG (Weis *et al.*, 1990), for H1N1pdm we use 4LXV (Yang *et al.*, 2014), for the influenza B Victoria and Yamagata lineages we use 4FQM (Dreyfus *et al.*, 2012) and 4M40 (Ni *et al.*, 2013).

References

Barr, I., C. Russell, T. Besselaar, N. Cox, R. Daniels, J. McCauley, T. Odagiri, S. Schultz-Cherry, Y. Shu, D. Smith, M. Tashiro, D. Wang, *et al.*, 2014, Vaccine **32**(37), 4713.

Barr, I. G., J. McCauley, N. Cox, R. Daniels, O. G. Engelhardt, K. Fukuda, G. Grohmann, A. Hay, A. Kelso, A. Klimov, T. Odagiri, D. Smith, *et al.*, 2010, Vaccine **28**(5), 1156, ISSN 0264-410X.

Bedford, T., M. A. Suchard, P. Lemey, G. Dudas, V. Gregory, A. J. Hay, J. W. McCauley, C. A. Russell, D. J. Smith, and A. Rambaut, 2014, eLife Sciences **3**, e01914, ISSN 2050-084X.

Candes, E., and T. Tao, 2005, IEEE Transactions on Information Theory **51**(12), 4203, ISSN 0018-9448.

Dreyfus, C., N. S. Laursen, T. Kwaks, D. Zuijdgeest, R. Khayat, D. C. Ekiert, J. H. Lee, Z. Metlagel, M. V. Bujny, M. Jongeneelen, R. van der Vlugt, M. Lamrani, *et al.*, 2012, Science **337**(6100), 1343, ISSN 0036-8075.

Gong, L. I., M. A. Suchard, and J. D. Bloom, 2013, eLife Sciences **2**, e00631, ISSN 2050-084X.

Harvey, W. T., V. Gregory, D. J. Benton, J. P. J. Hall, R. S. Daniels, T. Bedford, D. T. Haydon, A. J. Hay, J. W. McCauley, and R. Reeve, 2014, arXiv:1404.4197 [q-bio] ArXiv: 1404.4197.

Hay, A., R. S. Daniels, Y. P. Lin, X. Zheng, V. Gregory, M. Bennet, and L. Whittaker, 2008, *Characteristics of human influenza AH1N1, AH3N2, and B viruses isolated September 2007 to February 2008*, Technical Report, WHO Collaborating Centre for Reference and Research on Influenza, National Institute for Medical Research, UK.

Hay, A., R. S. Daniels, Y. P. Lin, X. Zheng, T. Hou, V. Gregory, L. Whittaker, J. Kloess, and N. Cattle, 2009, *Antigenic and genetic characteristics of human influenza A(H1N1), A(H3N2) and B viruses isolated during October 2008 to February 2009*, Technical Report, WHO Collaborating Centre for Reference and Research on Influenza, National Institute for Medical Research, UK.

Hay, A. J., V. Gregory, A. R. Douglas, and Y. P. Lin, 2001, Philos. Trans. R. Soc. Lond., B, Biol. Sci. **356**(1416), 1861, ISSN 0962-8436.

Hay, A. J., Y. P. Lin, V. Gregory, and M. Bennet, 2002, *Annual Report*, Technical Report, WHO Collaborating Centre for Refer-

ence and Research on Influenza, National Institute for Medical Research, UK.

He, J., and M. W. Deem, 2010, Protein Engineering, Design and Selection **23**(12), 935, ISSN 1741-0126, 1741-0134.

Hirst, G. K., 1942, J Exp Med **75**(1), 49, ISSN 0022-1007, 1540-9538.

JSmol developers, 2015, JSmol: an open-source html5 viewer for chemical structures in 3d., URL `http://wiki.jmol.org/index.php/JSmol`.

Koel, B. F., D. F. Burke, T. M. Bestebroer, S. v. d. Vliet, G. C. M. Zondag, G. Vervaet, E. Skepner, N. S. Lewis, M. I. J. Spronken, C. A. Russell, M. Y. Eropkin, A. C. Hurt, *et al.*, 2013, Science **342**(6161), 976.

Łuksza, M., and M. Lässig, 2014, Nature **507**(7490), 57.

McCauley, J., R. Daniels, Y. P. Lin, X. Zheng, V. Gregory, L. Whittaker, N. Cattle, C. Halai, and K. Cross, 2011, *Report prepared for the WHO annual consultation on the composition of influenza vaccine for the Southern Hemisphere*, Technical Report, WHO Collaborating Centre for Reference and Research on Influenza, National Institute for Medical Research, UK.

McCauley, J., R. Daniels, Y. P. Lin, X. Zheng, T. Hou, V. Gregory, L. Whittaker, N. Cattle, C. Halai, K. Cross, and J. Kloess, 2010a, *Report prepared for the WHO annual consultation on the composition of influenza vaccine for the Southern Hemisphere*, Technical Report, WHO Collaborating Centre for Reference and Research on Influenza, National Institute for Medical Research, UK.

McCauley, J., R. Daniels, Y. P. Lin, X. Zheng, T. Hou, V. Gregory, L. Whittaker, N. Cattle, J. Kloess, and C. Halai, 2010b, *Report prepared for the WHO annual consultation on the composition of influenza vaccine for the Northern Hemisphere*, Technical Report, WHO Collaborating Centre for Reference and Research on Influenza, National Institute for Medical Research, UK.

McCauley, J., R. Daniels, L. YP, Z. X, G. V, L. Whittaker, N. Cattle, C. Halai, and K. Cross, 2012a, Report prepared for the who annual consultation on the composition of influenza vaccine for the southern hemisphere, URL `http://www.crick.ac.uk/media/221907/interim-report-september-2012_2.pdf`.

McCauley, J., R. Daniels, L. YP, Z. X, G. V, L. Whittaker, N. Cattle, C. Halai, and K. Cross, 2013a, Report prepared for the who annual consultation on the composition of influenza vaccine for the northern hemisphere, URL `http://www.crick.ac.uk/media/221907/interim-report-february_2013.pdf`.

McCauley, J., R. Daniels, L. YP, Z. X, G. V, L. Whittaker, N. Cattle, C. Halai, K. Cross, and A. Rattigan, 2013b, Report prepared for the who annual consultation on the composition of influenza vaccine for the southern hemisphere, URL `http://www.crick.ac.uk/media/221907/interim-report-sep2013final.pdf`.

McCauley, J., R. Daniels, L. YP, Z. X, G. V, L. Whittaker, N. Cattle, C. Halai, K. Cross, A. Rattigan, and B. Ermetal, 2014a, Report prepared for the who annual consulta-

tion on the composition of influenza vaccine for the northern hemisphere, URL `http://www.crick.ac.uk/media/221907/interim-report-feb2014web.pdf`.

McCauley, J., R. Daniels, L. YP, Z. X, G. V, L. Whittaker, N. Cattle, C. Halai, K. Cross, A. Rattigan, and B. Ermetal, 2014b, Report prepared for the who annual consultation on the composition of influenza vaccine for the southern hemisphere, URL `http://www.crick.ac.uk/media/221907/interim-report-sep-14-web.pdf`.

McCauley, J., R. Daniels, L. YP, Z. X, G. V, L. Whittaker, N. Cattle, C. Halai, K. Cross, A. Rattigan, B. Ermetal, and M. Dai, 2015, Report prepared for the who annual consultation on the composition of influenza vaccine for the northern hemisphere, URL `http://www.crick.ac.uk/media/221907/interim-report-feb15-web.pdf`.

McCauley, J., R. S. Daniels, Y. P. Lin, X. Zheng, V. Gregory, L. Whittaker, N. Cattle, C. Halai, and K. Cross, 2012b, *Report prepared for the WHO annual consultation on the composition of influenza vaccine for the Northern Hemisphere*, Technical Report, WHO Collaborating Centre for Reference and Research on Influenza, National Institute for Medical Research, UK.

Muoz, E. T., and M. W. Deem, 2005, Vaccine **23**(9), 1144, ISSN 0264-410X.

Neher, R. A., and T. Bedford, 2015, Bioinformatics , btv381ISSN 1367-4803, 1460-2059.

Neher, R. A., C. A. Russell, and B. I. Shraiman, 2014, eLife Sciences **3**, e03568.

Ni, F., E. Kondrashkina, and Q. Wang, 2013, Virology **446**(1-2), 112, ISSN 1096-0341.

Russell, C. A., T. C. Jones, I. G. Barr, N. J. Cox, R. J. Garten, V. Gregory, I. D. Gust, A. W. Hampson, A. J. Hay, A. C. Hurt, J. C. d. Jong, A. Kelso, *et al.*, 2008, Science **320**(5874), 340, ISSN 0036-8075, 1095-9203.

Shih, A. C.-C., T.-C. Hsiao, M.-S. Ho, and W.-H. Li, 2007, Proc Natl Acad Sci USA **104**(15), 6283.

Smith, D. J., A. S. Lapedes, J. C. d. Jong, T. M. Bestebroer, G. F. Rimmelzwaan, A. D. M. E. Osterhaus, and R. A. M. Fouchier, 2004, Science **305**(5682), 371, ISSN 0036-8075, 1095-9203.

Steinbrück, L., T. R. Klingen, and A. C. McHardy, 2014, J. Virol. , JVI.01861ISSN 0022-538X, 1098-5514.

Steinbrück, L., and A. C. McHardy, 2011, Nucleic Acids Res **39**, e4.

Sun, H., J. Yang, T. Zhang, L.-P. Long, K. Jia, G. Yang, R. J. Webby, and X.-F. Wan, 2013, mBio **4**(4), e00230, ISSN , 2150-7511.

Thyagarajan, B., and J. D. Bloom, 2014, eLife Sciences **3**, e03300, ISSN 2050-084X.

Weis, W. I., A. T. Brünger, J. J. Skehel, and D. C. Wiley, 1990, J. Mol. Biol. **212**(4), 737, ISSN 0022-2836.

Yang, H., J. C. Chang, Z. Guo, P. J. Carney, D. A. Shore, R. O. Donis, N. J. Cox, J. M. Villanueva, A. I. Klimov, and J. Stevens, 2014, J. Virol. **88**(9), 4828, ISSN 1098-5514.

substitutions	85 to 95	90 to 00	95 to 05	00 to 10	05 to 16
K189N	—	—	—	3.44	2.65
K158N/N189K	—	—	—	2.17	3.25
K156E	—	3.18	—	—	—
C1*	—	3.03	—	—	—
S262N	—	2.61	0.0	—	0.04
K135G	1.9	2.17	—	—	—
C1R*	—	2.1	—	—	—
K158R	—	—	—	1.91	2.08
C2*	—	—	1.94	—	—
N145K	1.88	0.01	0.42	0.5	—
K193N	1.67	—	—	—	—
K140E	—	—	1.65	1.41	—
H155Y/R189K	1.57	—	—	—	—
T135G	—	1.55	—	—	—
K145N	0.47	0.0	1.53	1.18	—
S133D/E156K	—	1.49	—	—	—
S157L	0.47	0.79	—	—	1.4
Y155H/K189R	1.4	—	—	—	—
K135T	—	1.37	—	—	—
S193F/D225N	—	—	—	1.35	—
T212A	—	—	—	1.27	0.62
C3*	1.24	—	—	—	—
N144K	—	—	—	1.24	0.0
K144D	—	—	—	—	1.23
N121T	—	—	1.21	—	—
K62E/N276K	—	—	1.19	—	—
S159Y	1.17	—	—	—	0.83
S133D	1.17	—	—	—	—
S193N	1.16	—	0.56	0.0	—
L226V	—	0.19	1.13	—	—
S189R	—	1.1	—	—	—
G186S	—	—	1.07	—	—
E190D	0.0	1.06	—	—	—
K135E	1.04	1.05	—	—	—
Q156H	—	—	1.04	0.67	—
K144N	—	—	—	—	1.03
T135K	—	0.0	1.01	—	—
Y159F	—	—	1.01	0.9	0.0
I112V/S193F	—	—	—	—	1.01
T212S	—	—	—	—	1.0
K140I	—	—	—	0.74	1.0
K135D	—	0.99	—	—	—
K156E/D190E	0.97	—	—	—	—
F159Y	—	—	0.17	0.34	0.97
C4*	—	—	0.97	0.27	—
D172G/Q197R/N278S	0.96	—	—	—	—
T214I	0.11	0.91	—	—	—
R189S	0.9	0.7	—	—	—
K156H	—	—	0.9	—	—
L157S/S189R	0.89	—	—	—	—

TABLE S1 The 50 largest inferred antigenic effects of substitutions in HA1 for A/H3N2 in the past 30 years as inferred by the substitution model in overlapping 10 year intervals. When several substitutions always occurred together, a combined effect is shown. A dash indicated the absence of the substitution in a particular time interval. The substitutions are sorted by the maximum across time intervals. * C1: E62K, N121T, S124G, N133D, R142G, I144V, Q156K, K158E, A196V, K276N; C1R: K62E, V144I, K156Q, E158K, V196A, N276K, i.e., largely the reverse of cluster C1; C2: S124G, N133D, I144V, Q156K, K276N a subset of cluster C1; C3: K82E, E83K, A131T, R299K; C4: I25L, Q75H, T131A, T155H.
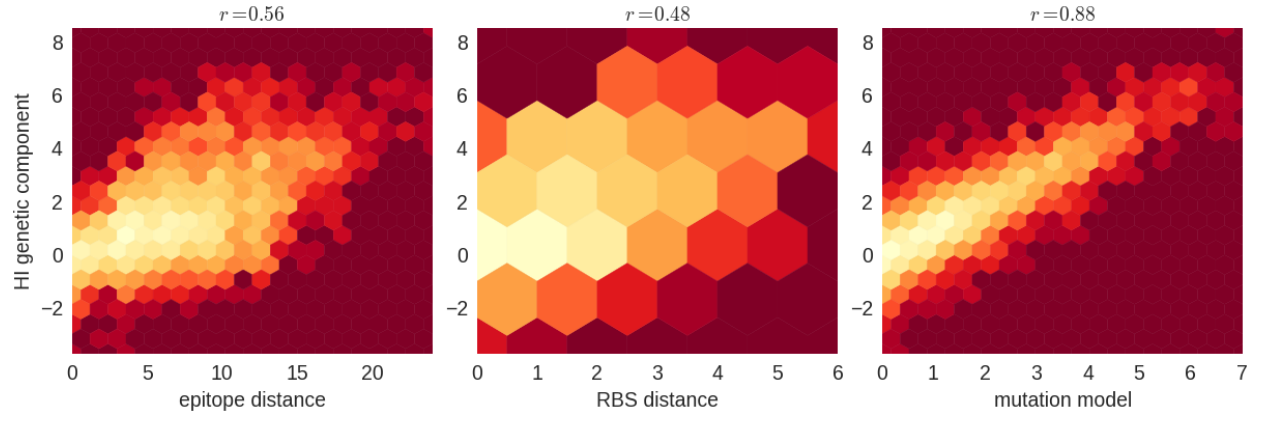
.

FIG. S1 Correlation of HI titer measurements (y-axis) with epitope distance between test and reference virus, distance at receptor binding sites, and the genetic component of the substitution model. Titer measurements are corrected by antiserum potency and virus avidity. Neither amino acid, epitope, or RBS distance explain much of the titer variation.
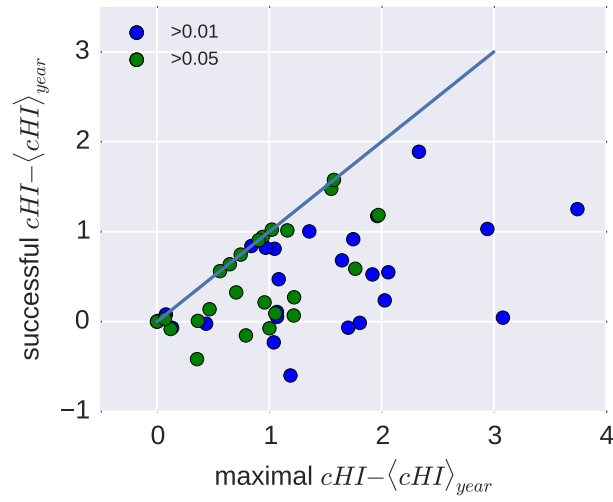

FIG. S2 The in many years, the most antigenically advanced clade is not the clade dominating the next season. The figure shows the maximally observed centered cHI vs the centered cHI of the successful clade for each year from 1990 to 2014.
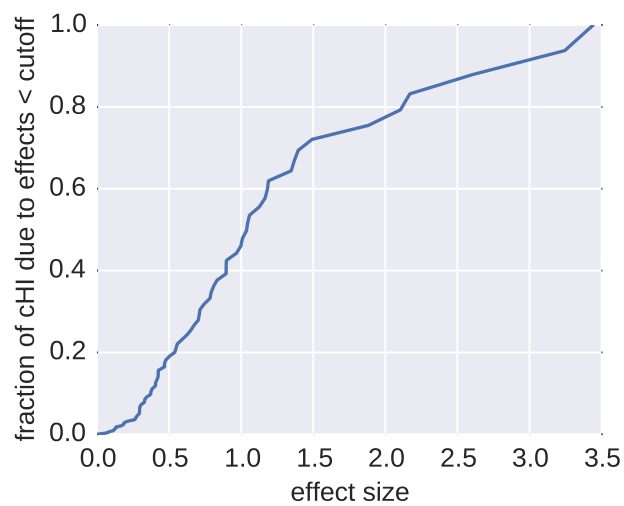
FIG. S3 About half of antigenic change is attributed to mutations with effects smaller than one unit. The figure shows the fraction antigenic change by common mutations with effects smaller than the cutoff on the horizontal axis.