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

Richard A. Neher[a], Trevor Bedford[b], Rodney S. Daniels[c], Colin A. Russell[d], and Boris I. Shraiman[e,1]

[a]Evolutionary Dynamics and Biophysics, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany; [b]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109; [c]Worldwide Influenza Centre, The Francis Crick Institute, London NW7 1AA, United Kingdom; [d]Department of Veterinary Medicine, University of Cambridge, Cambridge CB3 0ES, United Kingdom; and [e]Kavli Institute for Theoretical Physics, University of California, Santa Barbara, CA 93106

Human seasonal influenza viruses evolve rapidly, enabling the virus population to evade immunity and reinfect 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 A(H3N2) 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.

evolution | antigenic distance | phylogenetic tree

Seasonal influenza viruses evade immunity in the human population through frequent amino acid substitutions in their hemagglutinin (HA) and neuraminidase (NA) surface glycoproteins (1). 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) (2).

Antigenic properties of influenza viruses are measured in hemagglutination inhibition (HI) assays (3) 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. 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 twofold 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) test viruses. HI titers typically decrease with increasing genetic distance between reference and test viruses (1).

Given multiple antisera raised against different reference viruses and a panel of test viruses, WHO CCs routinely measure the HI titers $T_{a\beta}$ of all combinations of test viruses a and sera β, resulting in a matrix of HI titers (see Fig. 1A). The HI titer of a test virus a using antiserum β raised against the reference virus b is typically standardized as $H_{a\beta} = \log_2(T_{b\beta}) - \log_2(T_{a\beta})$, i.e., the difference in the number of twofold dilutions between homologous and heterologous titer. Standardized $\log_2$ titers from many HI assays can be visualized in two dimensions via multidimensional scaling—an approach termed "antigenic cartography" (4). Although standard cartography does not use sequence information,

sequences have been used as priors for positions in a Bayesian version of multidimensional scaling (5). To infer contributions of individual amino acid substitutions to antigenic evolution, Harvey et al. and Sun et al. (6, 7) have used models that predict HI titer differences by comparing sequences of reference and test viruses.

Here, we show that antigenic properties of seasonal influenza viruses are accurately described by a model based on the phylogenetic tree structure of their HA sequences. We use the model to show that HI titers have a largely symmetric and tree-like structure that can be used to define an antigenic distance between viruses. We show that large-effect substitutions account for about half of the total antigenic change and that the effect of specific substitutions is 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.org—an interactive real-time tracking tool for influenza virus evolution (8).

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 (9–13).

Results

We use two related models that predict HI titers from sequences. The first—the tree model—explains standardized $\log_2$ HI titers $H_{a\beta}$ as a sum of contributions associated with internal branches

Significance

Humans mount an antibody-mediated immune response against influenza viruses that can be recalled. Nevertheless, individuals can suffer from recurrent influenza infections as viruses can change their antigenic properties by altering their surface glycoproteins. This antigenic evolution requires frequent update of seasonal influenza vaccines. To inform vaccine updates, laboratories that contribute to the World Health Organization Global Influenza Surveillance and Response System assess the antigenic phenotypes of circulating viruses. Based on the relationship of antigenic distance to genetic differences between viruses, we developed a model to interpret measured antigenic data and predict the properties of viruses that have not been characterized antigenically and explore the model's value in predicting the future composition of influenza virus populations.
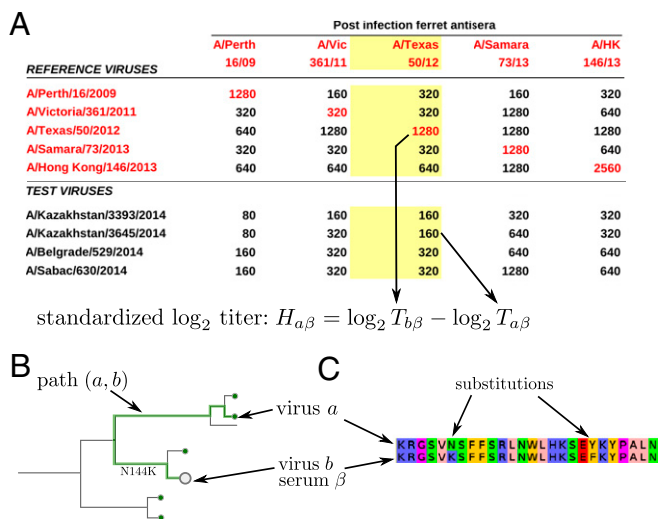
Fig. 1. Antigenic data and models for HI titers. (*A*) A typical table reporting HI titer data. Each number in the table is the maximum dilution at which the antiserum (column) inhibited hemagglutination of red blood cells by a virus (row). The red numbers on the diagonal indicate homologous titers. 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. To make measurements using different antisera comparable, we define standardized log-transformed titers $H_{a\beta}$ relative to the homologous titer. (*B*) Each HI titer between antiserum α and virus b can be associated with a path on the tree connecting the reference and test viruses a and b, respectively, indicated as a thick line. The tree model seeks to explain the antigenic differences 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.

in the phylogenetic tree that connect the reference virus b and the test virus a (see Fig. 1*B*). The second—the substitution model—explains HI titers in terms of a sum of contributions associated with amino acid substitutions between reference and test virus (see Fig. 1*C*). These two models are similar but complement each other in a few aspects that we discuss further below. In addition to contributions associated with branches or substitutions, measured HI titers depend on the overall reactivity of individual antisera and viruses 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 measured standardized $\log_2$ titer $H_{a\beta}$ between virus a and antiserum β (corresponding to the reference virus b) is modeled by $\hat{H}_{a\beta}$ defined as

$$H_{a\beta} \approx \hat{H}_{a\beta} = v_a + p_\beta + D_{ab} \qquad [1]$$

where v_a, p_β, and D_{ab} denote the avidity of virus a, the potency of antiserum β, and the genetic component of HI titer differences, respectively. The latter is decomposed into a sum of contributions d_i along internal branches i in the path $(a \ldots b)$ separating the test (a) and reference viruses (b), i.e., $D_{ab} = \sum_{i \in (a \ldots b)} d_i$. In the substitution model, the sum over branches is replaced by a sum over amino acid substitutions between reference and test viruses. The parameters of our model are determined by fitting the model to available HI titer measurements while penalizing large values for the parameters.

The virus avidity v_a accounts for systematic variations of HI titers of virus a across multiple antisera, i.e., a row of the HI matrix in Fig. 1. Within our model, v_a can be positive or negative.

Large absolute values of v_a are penalized by adding a term proportional to v_a^2 to the cost function (ℓ_2-norm regularization). Similarly, the antiserum potency p_β captures variation in HI titers of antiserum β across many test viruses, i.e., a column of the HI matrix. Part of the latter variation is already removed by using standardized titers $H_{a\beta}$ relative to the homologous titer $T_{b\beta}$, but systematic variation remains that is absorbed by p_β. Potencies are regularized by their ℓ_2-norm, as well, and allowed to be positive or negative.

Titers tend to decrease with increasing genetic distance between the reference and test virus. We therefore constrain the contributions d_i to be nonnegative. While similar, the tree and substitution models differ slightly in how the genetic component of HI titers is parameterized. The tree model associates one term d_i with each branch, and the contribution of the branch is independent of the direction of the path running through the branch. The substitution model associates a nonnegative effect with each amino acid difference—D_{ab} is modeled as a weighted sum of amino acid differences between reference virus b and test virus a. We give the model the additional freedom of independent effects for forward and backward substitutions, e.g., F159S vs. S159F.

In contrast to the potencies and avidities, we regularize the d_i using their ℓ_1-norm, i.e., the d_i contribute to cost function via their absolute value rather than their square. This regularization encourages a sparse model in which a minority of d_i explain most antigenic evolution while many $d_i = 0$. Only internal branches of the tree or substitutions observed in more than one virus are candidates for nonzero contributions. Contributions of terminal branches or singletons can be absorbed into the virus avidities. For a detailed description of the models and inference procedures, see *Material and Methods*.

The Tree and Substitution Models Accurately Predict HI Titers. We evaluated the performance of the models in predicting HI titers for different influenza lineages A(H3N2), A(H1N1)pdm09, B/Yam, and B/Vic. We trained the models on 90% of the data and used the remaining measurements to validate the models as in ref. 5. The number of viruses, number of antisera, HI measurements, etc., for each lineage are provided as Table S1.

We found that the models were able to predict titers of antiserum−virus combinations to an accuracy of approximately $0.5 \log_2$ titer levels for A(H3N2) with somewhat lower accuracy for the influenza B lineages (Table 1 and Fig. 2*A*).

To quantify the prediction accuracy for viruses for which no antigenic data exist, 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 (Table 1 and Fig. 2*B*). Having completely excluded a virus from the training data implies that no avidity of a virus can be estimated (and $v_a = 0$ is used instead). The increased prediction error is therefore largely due to virus-to-virus variability that is not captured by the HA phylogeny.

Table 1. Prediction accuracy of the tree and substitution models across influenza viruses

Model	Tree		Substitution	
Untrained unit	Titer	Virus	Titer	Virus
A(H3N2) (12 y)	0.52	0.66	0.5	0.73
A(H1N1)pdm09 (7 y)	0.44	0.72	0.45	0.9
B/Yam (12 y)	0.64	0.88	0.65	1.0
B/Vic (12 y)	0.77	0.8	0.73	0.84

"Titer" and "virus" in the row "Untrained unit" refer to predictions for test sets in which individual titer measurements or all measurements involving a particular virus, respectively, were omitted from the training data.
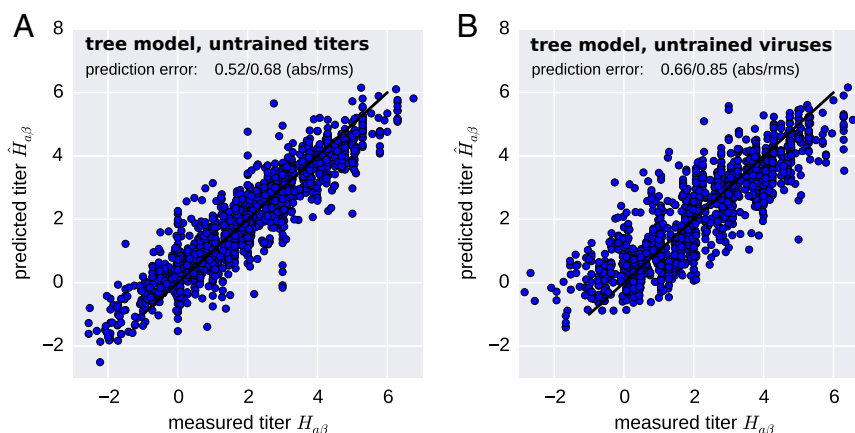
Fig. 2. A(H3N2) HI titers are accurately predicted by the tree model. The figure scatters predicted titers $\hat{H}_{a\beta}$ (y axis) against a test set of measurements $H_{a\beta}$ not used for training of the model (x axis). This test set either consists of (A) a random sample of 10% of all measurements or (B) all measurements for 10% of all viruses. In the latter, no avidity can be estimated for viruses in the test set because these viruses are completely absent from the training data ($v_a = 0$ is assumed). Hence prediction accuracy is lower but still comparable to the measurement accuracy. The figure shows data obtained using a 12-y set of A(H3N2) viruses. Other time intervals and lineages are predicted similarly well.

To infer the genetic component D_{ab} 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). Because the substitution or branch effects pick up antigenic changes associated with a larger number of antiserum–virus pairs, whereas antiserum potencies and virus avidities absorb serum- and virus-specific variation, the resulting model of antigenic distances D_{ab} is a smoothed and coarse-grained description of the HI titer data.

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.

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, cHI, for each node in the tree. 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). Our models infer that A (H3N2) viruses advance by ~0.7 log$_2$ titer units per year (Fig. 3A), whereas influenza B virus lineages advance 0.15 (Vic) and 0.1 (Yam) units per year (Fig. 3B). Previous estimates of these rates using cartography (5) suggested higher absolute values, but the relative magnitude of rates in different lineages were found to be similar. Forcing the antigenic distance matrix into two dimensions causes distortions that might result in an exaggeration of the distances. Very little consistent temporal change in antigenic properties is observed in A(H1N1)pdm09 (Fig. 3B).

We estimate that, for A(H3N2) viruses, approximately half the cumulative antigenic evolution is due to a large number of substitutions with effects smaller than one unit, whereas 20% is accounted for by a few substitutions with effects greater than two units (Fig. S1). However, some of the largest effects are associated with clusters of colinear amino acid substitutions, and their individual effects cannot be resolved.
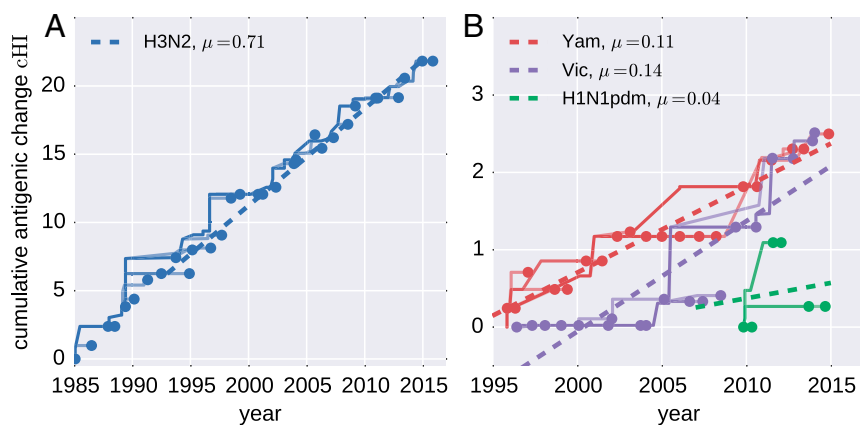
Fig. 3. Cumulative antigenic change (cHI) from phylogeny root to tips across influenza viruses. (A) A(H3N2) antigenic evolution over the past 30 y. The trunk of the tree is shown as a thick line, and side branches are shown as thin lines. The dashed straight line indicates the linear regression vs. time. (B) The corresponding traces for the two influenza B lineages (past 20 y) and A(H1N1)pdm09 (past 7 y).

As we have shown, the two models of HI titers accurately predict HI titers of all seasonal influenza virus lineages. However, A(H3N2) viruses have been monitored more closely and over a larger number of years than the other lineages. We therefore focus subsequent analysis of the structure of antigenic space and the dynamics of antigenic evolution on A(H3N2).

HI Titers Define an Approximate Distance. To further evaluate the approach of partitioning a titer into an antiserum, a virus, and a symmetric tree component D_{ab}, 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 $\Delta_{a\beta} = H_{a\beta} - v_a - p_\beta$ and $\Delta_{b\alpha}$ should reduce the titers to the symmetric tree component $D_{ab} = D_{ba}$. Fig. 4A compares the distribution of $\Delta_{a\beta} - \Delta_{b\alpha}$ with the uncorrected difference between the reciprocal titers $H_{a,\beta} - H_{b,\alpha}$ for A(H3N2). Although raw reciprocal titer measurements often differ by several $\log_2$ titer units (SD 2.0), the corrected tree component was symmetric to within one unit (SD 0.9), comparable to the accuracy of HI titer measurements.

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. The quartet rule can only be tested on sets of viruses for which all pairwise distances have been inferred. Hence, we determined maximal "cliques" of reference antisera, the activities of which had been measured against all other reference viruses within the clique. Out of these cliques, we repeatedly selected, at random, four viruses, calculated the three distance sums, and compared the largest and second largest distance sum. As a control, we constructed the analogous distribution for triplets of random variables with the same distribution as the distance sums, but without the dependence via the tree: We obtained three sums of two distances by three times randomly drawing four antisera. Both distributions are shown in Fig. 4B. The difference between the largest and second largest of these distance sums of quartets has a mean of 0.41 $\log_2$ titer levels, whereas the control has a

mean of 0.96 $\log_2$ titer levels. 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 (14, 15). Seven positions near the HA receptor binding site (Koel 7) have been shown to account for major historic antigenic transitions (16). However, the number of substitutions at epitope sites or Koel 7 sites alone is a poor predictor of antigenic distance, with r^2 values of about 0.25 (Fig. S2), possibly because the effect of a substitution 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. Table S2 shows the top contributions for A(H3N2) for sequences estimated from sets of sequences and the associated HI titers in five overlapping 10-y 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 are enough data to constrain the model.

By and large, the effects we infer are compatible with those associated with cluster transitions determined in ref. 16. For example, the transition from the Sichuan/87 (SI87) cluster to the Beijing/89 (BE89) cluster involved the substitution N145K, for which we infer an effect of 1.91 antigenic units (each unit is equal to a twofold 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 units 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, whereas the typical titer drop is ~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 ~3 units of which 2 units are attributed to the set of substitutions: K62E/V144I/ K156Q/E158K/V196A/N276K (C1). These substitutions account
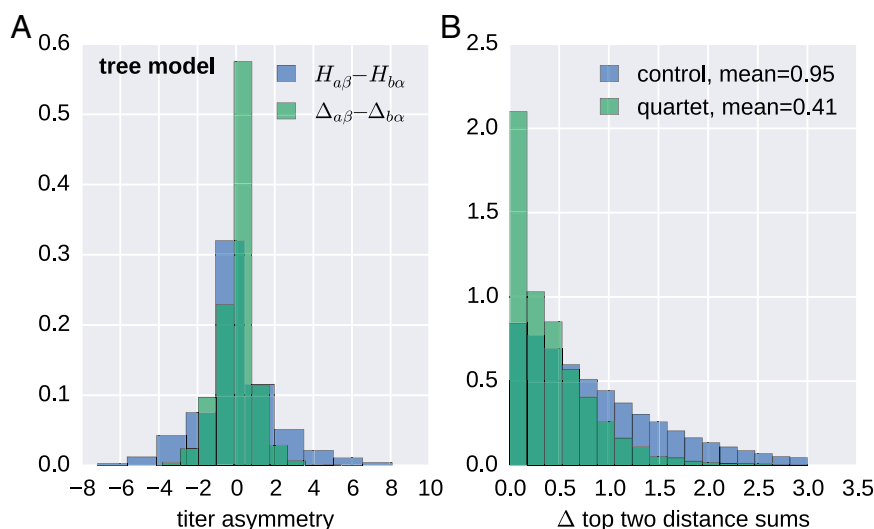
Fig. 4. H3N2 HI titers define an approximate a distance on the tree. (A) Reciprocal HI titer measurements often differ by several $\log_2$ units (histogram of $H_{a\beta} - H_{b\alpha}$). After subtracting avidities and potencies, the remainder $\Delta_{a\beta} - \Delta_{b\alpha}$ is almost symmetric, with deviations on the order of the measurement accuracy (SD 0.95). (B) On a tree, the two largest sums of distances within antisera/virus quartets are equal (see *HI Titers Define an Approximate Distance*). B 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. A and B both show results for A(H3N2) viruses from the last 12 years.

for 3.6 units in cartography. Lastly, the transition from SY97 to Fujian/02 (FU02) is attributed to Q156H (1.1 units as in ref. 16) 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 1989–1992 and K145N in 2004) is large in these instances, whereas the inferred effect for the backward direction is small. Some substitutions at Koel 7 positions have very small effects, such as, for example, S189N. 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-y intervals is given as Dataset S1 and the 55 largest effects are given in Table S2. However, the interpretation of effects is sometimes difficult due to colinearities between substitutions.

Limitations of the Models. The tree and substitution model perform similarly in terms of accuracy, as summarized in Table 1, which is expected because 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. Although 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. As a recent example, independent substitutions at HA1 position 159 of A(H3N2) viruses define different genetic/antigenic clades: F159Y defining clade 3C.2a and F159S defining clade 3C.3a. The distances between viruses in these clades are not necessarily a sum of the effects associated with the branches separating the clades, because one substitution at position 159 masks the other. 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 differentiates viruses in these clades.

Similarly, the substitution model fails when the same substitution (as opposed to different substitutions at the same position) occurred in different places on the tree in different genetic backgrounds, which is common in sequence ensembles covering long time periods. 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" measurements are underrepresented in the data, and although 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 models accurately predict 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 in *Visualization of Antigenic Evolution* 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. We investigated the relationship between the amount of antigenic change and success of clades in the phylogenetic tree. Fig. 5 shows frequency trajectories of clades reaching at least 10% at one time, and color indicates the magnitude of the antigenic change that accumulated along the ancestral lineage over the preceding 6 mo. Consistent with expectation, large antigenic changes fix more often than small antigenic changes. However, there are also several clades that evolved antigenically but failed to spread. Extinction of transiently successful clades could be due to fitness cost associated with the substitutions responsible for antigenic change or could be the result of competition of multiple antigenically advanced clades (17–19).

For each season, we determined the clade with highest Local Branching Index [LBI, a predictor of clade success (10)] and the clade with the largest antigenic advancement (cHI) relative to all other viruses in a season. We restricted the latter 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 to the future population of the most antigenically advanced clade is significantly below the population average (dashed line in Fig. 5B). Predictions by cHI and LBI are comparable in quality and are correlated. However, clades with maximal cHI are sometimes far from the future population, suggesting that a predictor based on antigenic phenotype alone readily generates false positives, i.e., highest scoring clades that go extinct; the problem becomes worse when smaller clades are included (threshold lowered to >1% of all sequences). Nevertheless, successful clades tend to be antigenically advanced. Fig. 5C shows the distribution of cHI of clades closest to the next season. These clades tend to have larger cHI than cocirculating viruses. These results suggest that cHI correlates with spread of a clade but is not the only factor determining clade success (see also Fig. S3).

As an alternative to clades in the tree, Fig. 6 shows the probability of a substitution reaching a certain frequency as a function of its inferred antigenic effect. The probability to rise to a high frequency increases with antigenic effect, but even substitutions with very large effect can fail to spread—an effect that limits predictive power. This occasional lack of fixation happens despite the fact that antigenic effects of successful substitutions are more likely to be detected.

Visualization of Antigenic Evolution. Antigenic evolution can be visualized by mapping titer distances into the plane using a variant of multidimensional scaling (4). This 2D representation, however, is not readily superimposed with the sequence evolution of the viruses. Instead of squeezing the sequence evolution onto the plane (5), we map the HI titer data onto the phylogenetic tree using nextflu (8).

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 nextflu.org/h3n2/3y/, 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 gray squares. A focal reference virus can be chosen by clicking on one of these squares, thereby coloring the tree according to the average antigenic distance between viruses and antisera raised against the focal
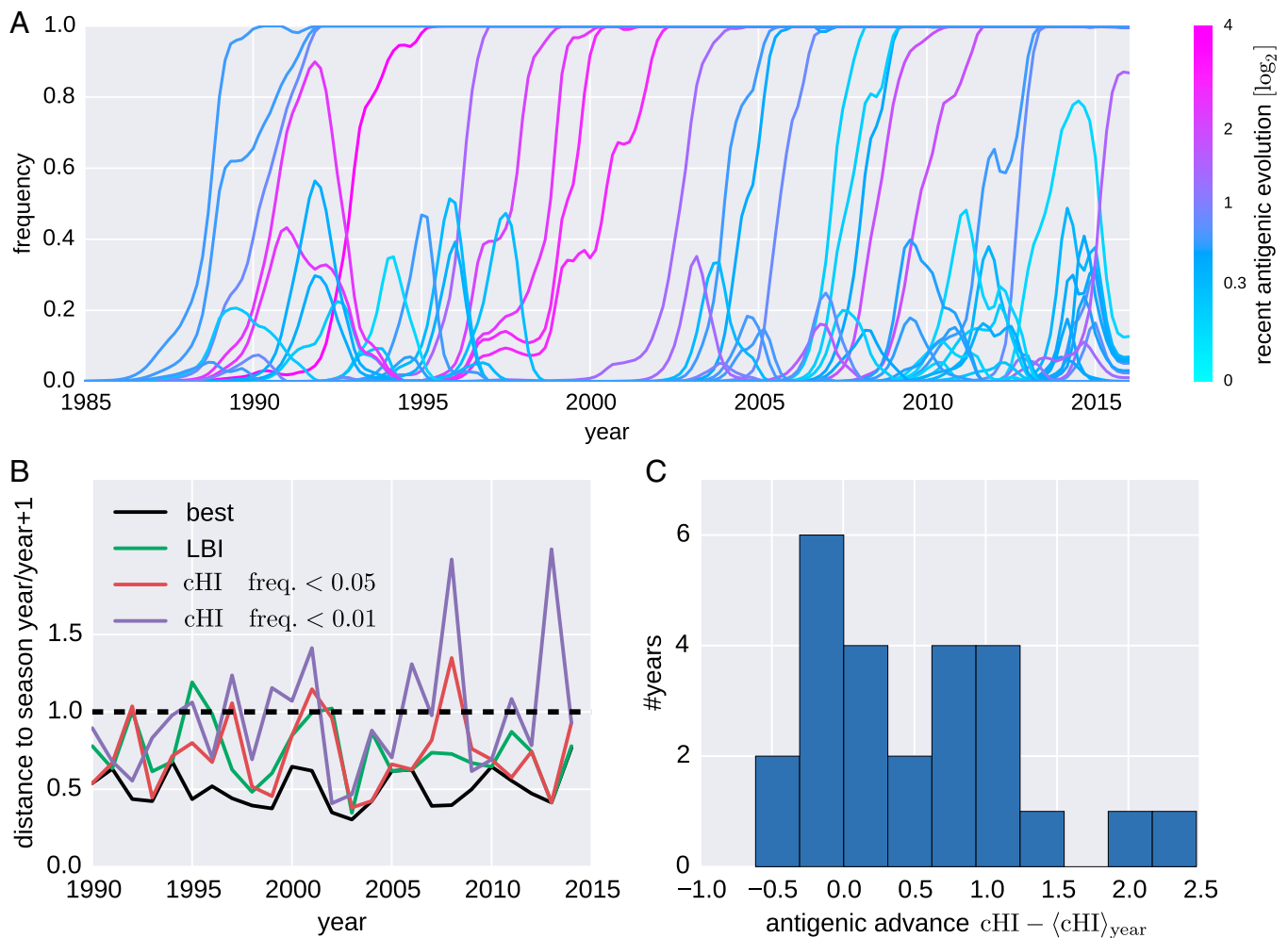
Fig. 5. Antigenic evolution and the success of A(H3N2) clades. (*A*) Trajectories of clades colored by inferred cumulative antigenic evolution in the past 6 mo on branches ancestral to the clade. The majority of the clades with substantial recent antigenic change fix, but several antigenically evolved branches go extinct. *B* quantifies the power of cumulative antigenic change (cHI) to predict clade success. *B* shows the nucleotide distances of clades with maximal cHI among clades that account for at least 1% and 5% of all viruses. Predictions based on maximal LBI are also shown. Distances are relative to the average distance between present and future populations shown as the dashed line at 1. The solid black line indicates the distances of the best possible pick in each year. Predictions based on maximal LBI and cHI are correlated, but maximal cHI is sensitive to false positives (highest scoring clades that go extinct). (*C*) Histogram of centered cHI of the clade closest to the next season for each year, suggesting that successful clades tend to be antigenically advanced.

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.

In addition to titer measurements and titer predictions, the website allows coloring of the tree according to 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 cartographic approaches (4). In addition to prediction of HI titers, the models and the mapping of HI data onto the HA phylogeny provide insight into the structure of antigenic space and allowed us to investigate the relationship between antigenic evolution and clade success.

Previous analyses had concluded that two dimensions provide the optimal embedding for antigenic evolution of human seasonal H3N2 viruses, and adding additional dimensions did not improve the predictive power (4, 5). Here we find that a model based on the tree structure—effectively infinite dimensional—predicts titers with similar or better accuracy. This apparent discrepancy has its roots in the number of parameters necessary to specify the model. A d-dimensional antigenic map of V viruses and S antisera has $d(V + S - 1)$ location parameters. If, in addition to locations, avidities and potencies are inferred, the parameter count increases by $V + S$. Because the number of parameters increases rapidly with the number of dimensions of the map, the predictive power can decrease at high d due to overfitting. In contrast, the tree model requires only one parameter for every internal branch in the tree (of which there are,
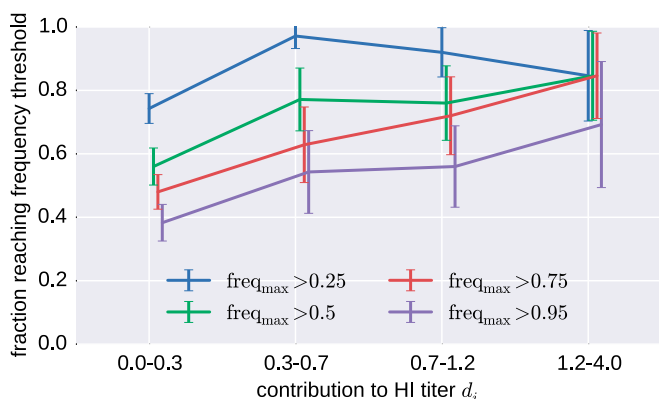
Fig. 6. Substitutions with large antigenic effect fix preferentially. The figure shows the fraction of all substitutions in A(H3N2) that reach the indicated frequency for different magnitudes of the inferred antigenic effect of this substitution. The plot combines data from substitution models fitted to five overlapping 10-y 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 SDs of boot strap replicates of substitutions.

at most, $B < V - 2$) and $V + S$ parameters for avidities and potencies. In practice, the number of internal branches B is substantially lower than $V - 2$ (often severalfold) due to the many polytomies in the tree. For A(H3N2) viruses from the past 12 y, HI data were available for 1,772 out of a total of 2,502 viruses in the tree. Of the 610 internal branches constrained by HI data, 231 branches were inferred to affect HI titers (with effects >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 "treeness" using quartet distances and symmetry between reciprocal measurements (Fig. 4).

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 ref. 5. 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 2D 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 (9–13) 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 (10). 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 results in the loss of 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 or (*ii*) by a fitness cost associated with a large antigenic effect substitutions. Competition between clades in terms of antigenic advancement and mutational load has been demonstrated in computational models (18) and is the basis of recent efforts to predict influenza virus evolution (9). 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 (17, 20) and have also been studied in computational models (19). Better understanding of the context dependence of large antigenic effect HA substitutions may therefore be a promising path toward reducing the false positive rate and improving prediction capacity. The problem of false positive detection is also seen at the Koel 7 positions: Although most past dramatic changes in antigenic phenotype are associated with substitutions at these positions, they do not always have antigenic effects and/or may fail to spread. Hence Koel 7 substitutions alone are poor predictors of clade success.

In conclusion, our study demonstrates that HI data integrate naturally onto the sequence-derived phylogeny of the virus. Although, at present, HI-based prediction does not outperform sequence-based methods, better understanding of genetic context dependence of HI data may provide a path toward 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.
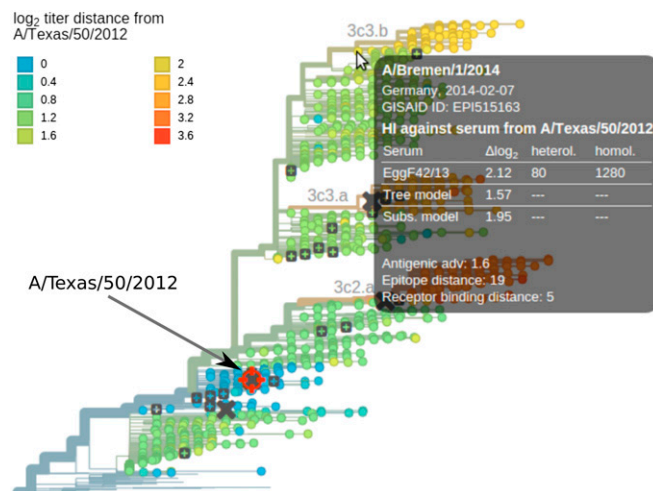
Fig. 7. Visualization of antigenic evolution on the tree. Tree tips are colored according to the predicted $\log_2$ titer distance relative the focal virus (A/Texas/50/2012 in this case). The tool tip shows titers relative to all available antisera and titer predictions as a table. Each reference virus (gray 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, and the color averages measurements relative to different antisera. Visualizations are available for all seasonal influenza lineages.

Materials and Methods

Data HA Sequences of Influenza. A and B viruses isolated from humans were downloaded from the Global Initiative on Sharing Avian Influenza Data (GISAID). Accession numbers of all sequences are provided as [Datasets S2–S5](#). We collected HI data from references (4, 21, 22) and annual and interim reports of the WHO CC London between 2002 and 2015 (23–35). HI data before 2011 were curated in ref. 5. Original HI data tables are available from the website of the Worldwide Influenza Centre at the Francis Crick Institute at https://www.crick.ac.uk/research/worldwide-influenza-centre/annual-and-interim-reports/.

Although the modalities of HI assays have changed over the years (red blood cells from different species, addition of NA 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, are modeled with greater accuracy.

Data Processing and Model Fit. The data processing pipeline is based on nextflu (8), which subsamples viruses, aligns sequences, and builds a phylogenetic tree. To make use of as much HI data as possible, viruses that are antigenically characterized were preferentially included. This pipeline was modified for the current purpose to enforce the inclusion of all strains for which antigenic data were 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 standardized log titer as $H_{a\beta} = \log_2(T_{a\beta}) - \log_2(T_{b\beta})$, where $T_{a\beta}$ is the titer of antiserum to reference virus b required to inhibit virus a and $T_{b\beta}$ is the homologous titer. In case multiple measurements are available, we average the standardized log titers. When no homologous titer is available, the maximal titer is used as a proxy for the homologous titer. The path between test virus a and reference virus b extends over branches of the phylogeny $i \in (a \ldots b)$, where each branch makes a contribution d_i. Our model of HI titers $\hat{H}_{a\beta}$ between viruses a and serum β is defined in Eq. **1**. The parameters d_i, p_α, and v_a are then estimated by minimizing the cost function

$$C = \sum_{a,\beta} \left(H_{a\beta} - \hat{H}_{a\beta}\right)^2 + \lambda \sum_i d_i + \gamma \sum_a v_a^2 + \delta \sum_\alpha p_\alpha^2 \tag{2}$$

subject to the constraints $d_i \geq 0$. 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 (36). 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 refs. 6 and 7 for a similar approach). Because 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 0.2; $\lambda = 1.0$ 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 numbers 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 nonzero branch effects, and the total number of nonzero parameters is not much larger than the number of test viruses. In the substitution model, the number of nonsingleton 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

$$\text{minimize:} \quad \frac{1}{2} \mathbf{x} \mathbf{Q} \mathbf{x} + \mathbf{q} \mathbf{x}$$
$$\text{subject to:} \quad \mathbf{G} \mathbf{x} \leq \mathbf{h} \tag{3}$$

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_i associated with internal branches i of the tree, the potencies p_α of each antiserum, and avidities v_a of virus isolates into a single vector $\mathbf{x}$,

$$x_s = \begin{cases} d_s, & \text{for } s = 1 \ldots B \\ v_s, & \text{for } s = B+1 \ldots B+V \\ p_s, & \text{for } s = B+V+1 \ldots B+V+S \end{cases} \tag{4}$$

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 $\hat{H}_{a\beta}$. Using the double index $a\beta$ to label measurements, entries of $\mathbf{A}$ are given by

$$\mathbf{A}_{a\beta,s} = \begin{cases} 1, & \text{for } s \text{ in path, potency, or avidity} \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

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

$$\hat{H} = \mathbf{A} \mathbf{x} \tag{6}$$

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

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

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

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

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

we have

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

To enforce 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

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

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's visualization software auspice. 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 (37). For H3N2, we use structure 5HMG (38); for H1N1pdm, we use 4LXV (39); and, for the influenza B Victoria and Yamagata lineages, we use 4FQM (40) and 4M40 (41). The structure visualization is available at hi.nextflu.org/H3N2/3y/.

1. Hay AJ, Gregory V, Douglas AR, Lin YP (2001) The evolution of human influenza viruses. *Philos Trans R Soc Lond B Biol Sci* 356(1416):1861–1870.
2. Barr IG, et al.; Writing Committee of the World Health Organization Consultation on Northern Hemisphere Influenza Vaccine Composition for 2013–2014 (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(37):4713–4725.
3. Hirst GK (1942) The quantitative determination of influenza virus and antibodies by means of red cell agglutination. *J Exp Med* 75(1):49–64.
4. Smith DJ, et al. (2004) Mapping the antigenic and genetic evolution of influenza virus. *Science* 305(5682):371–376.
5. Bedford T, et al. (2014) Integrating influenza antigenic dynamics with molecular evolution. *eLife* 3:e01914.
6. Harvey WT, et al. (2014) Identifying the genetic basis of antigenic change in influenza A(H1n1). arXiv:1404.4197.
7. Sun H, et al. (2013) Using sequence data to infer the antigenicity of influenza virus. *MBio* 4(4):e00230—13.
8. Neher RA, Bedford T (2015) nextflu: Real-time tracking of seasonal influenza virus evolution in humans. *Bioinformatics* 31(21):3546–3548.
9. Łuksza M, Lässig M (2014) A predictive fitness model for influenza. *Nature* 507(7490):57–61.
10. Neher RA, Russell CA, Shraiman BI (2014) Predicting evolution from the shape of genealogical trees. *eLife* 3:e03568.
11. Steinbrück L, McHardy AC (2011) Allele dynamics plots for the study of evolutionary dynamics in viral populations. *Nucleic Acids Res* 39(1):e4.
12. He J, Deem MW (2010) Low-dimensional clustering detects incipient dominant influenza strain clusters. *Protein Eng Des Sel* 23(12):935–946.
13. Steinbrück L, Klingen TR, McHardy AC (2014) Computational prediction of vaccine strains for human influenza A (H3n2) viruses. *J Virol* 88(20):12123–12132.
14. Shih ACC, Hsiao TC, Ho MS, Li WH (2007) Simultaneous amino acid substitutions at antigenic sites drive influenza A hemagglutinin evolution. *Proc Natl Acad Sci USA* 104(15):6283–6288.
15. Muñoz ET, Deem MW (2005) Epitope analysis for influenza vaccine design. *Vaccine* 23(9):1144–1148.
16. Koel BF, et al. (2013) Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution. *Science* 342(6161):976–979.
17. Gong LI, Suchard MA, Bloom JD (2013) Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife* 2:e00631.
18. Koelle K, Rasmussen DA (2015) The effects of a deleterious mutation load on patterns of influenza A/H3n2's antigenic evolution in humans. *eLife* 4:e07361.
19. Kucharski A, Gog JR (2012) Influenza emergence in the face of evolutionary constraints. *Proc R Soc Lond B Biol Sci* 279(1729):645–652.
20. Thyagarajan B, Bloom JD (2014) The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. *eLife* 3:e03300.
21. Russell CA, et al. (2008) The global circulation of seasonal influenza A (H3N2) viruses. *Science* 320(5874):340–346.
22. Barr IG, et al.; Writing Committee of the World Health Organization Consultation on Northern Hemisphere Influenza Vaccine Composition for 2009-2010 (2010) Epidemiological, antigenic and genetic characteristics of seasonal influenza A(H1N1), A(H3N2) and B influenza viruses: basis for the WHO recommendation on the composition of influenza vaccines for use in the 2009-2010 northern hemisphere season. *Vaccine* 28(5):1156–1167.
23. Hay AJ, Lin YP, Gregory V, Bennet M (2002) *Annual Report* (Natl Inst Med Res, London).
24. Hay A, et al. (2008) *Characteristics of Human Influenza AH1N1, AH3N2, and B Viruses Isolated September 2007 to February 2008* (Natl Inst Med Res, London).
25. Hay A, et al. (2009) *Antigenic and Genetic Characteristics of Human Influenza A (H1N1), A(H3N2) and B Viruses Isolated During October 2008 to February 2009* (Natl Inst Med Res, London).
26. McCauley J, et al. (2010) *Report Prepared for the WHO Annual Consultation on the Composition of Influenza Vaccine for the Northern Hemisphere* (Natl Inst Med Res, London).
27. McCauley J, et al. (2010) *Report Prepared for the WHO Annual Consultation on the Composition of Influenza Vaccine for the Southern Hemisphere* (Natl Inst Med Res, London).
28. McCauley J, et al. (2011) *Report Prepared for the WHO Annual Consultation on the Composition of Influenza Vaccine for the Southern Hemisphere* (Natl Inst Med Res, London).
29. McCauley J, et al. (2012) *Report Prepared for the WHO Annual Consultation on the Composition of Influenza Vaccine for the Northern Hemisphere* (Natl Inst Med Res, London).
30. McCauley J, et al. (2012) *Report Prepared for the WHO Annual Consultation on the Composition of Influenza Vaccine for the Southern Hemisphere* (Natl Inst Med Res, London).
31. McCauley J, et al. (2013) *Report Prepared for the WHO Annual Consultation on the Composition of Influenza Vaccine for the Northern Hemisphere* (Natl Inst Med Res, London).
32. McCauley J, et al. (2013) *Report Prepared for the WHO Annual Consultation on the Composition of Influenza Vaccine for the Southern Hemisphere* (Natl Inst Med Res, London).
33. McCauley J, et al. (2014) *Report Prepared for the WHO Annual Consultation on the Composition of Influenza Vaccine for the Northern Hemisphere* (Natl Inst Med Res, London).
34. McCauley J, et al. (2014) *Report Prepared for the WHO Annual Consultation on the Composition of Influenza Vaccine for the Southern Hemisphere* (Natl Inst Med Res, London).
35. McCauley J, et al. (2015) *Report Prepared for the WHO Annual Consultation on the Composition of Influenza Vaccine for the Northern Hemisphere* (Natl Inst Med Res, London).
36. Candes E, Tao T (2005) Decoding by linear programming. *IEEE Trans Inf Theory* 51(12):4203–4215.
37. JSmol Developers (2015) JSmol: An open-source HTML5 viewer for chemical structures in 3D. Available at https://sourceforge.net/projects/jsmol. Accessed March 8, 2015.
38. Weis WI, Brünger AT, Skehel JJ, Wiley DC (1990) Refinement of the influenza virus hemagglutinin by simulated annealing. *J Mol Biol* 212(4):737–761.
39. Yang H, et al. (2014) Structural stability of influenza A(H1N1)pdm09 virus hemagglutinins. *J Virol* 88(9):4828–4838.
40. Dreyfus C, et al. (2012) Highly conserved protective epitopes on influenza B viruses. *Science* 337(6100):1343–1348.
41. Ni F, Kondrashkina E, Wang Q (2013) Structural basis for the divergent evolution of influenza B virus hemagglutinin. *Virology* 446(1-2):112–122.