Science

AAAS

Mapping the Antigenic and Genetic Evolution of Influenza Virus
Derek J. Smith *et al.*
Science **305**, 371 (2004);
DOI: 10.1126/science.1097211

CO_2 from preindustrial levels will result in a 30% decrease in carbonate ion concentration and a 60% increase in hydrogen ion concentration. As the carbonate ion concentration decreases, the Revelle factor increases and the ocean's ability to absorb more CO_2 from the atmosphere is diminished. The impact of this acidification can already be observed today and could have ramifications for the biological feedbacks in the future (26). If indeed the net feedbacks are primarily positive, the required socioeconomic strategies to stabilize CO_2 in the future will be much more stringent than in the absence of such feedbacks. Future studies of the carbon system in the oceans should be designed to identify and quantitatively assess these feedback mechanisms to provide input to models that will determine the ocean's future role as a sink for anthropogenic CO_2.

References and Notes

1. P. J. Crutzen, E. F. Stoermer, *Global Change Newslett.* **41**, 12 (2000).
2. R. A. Houghton, J. L. Hackler, in *Trends: A Compendium of Data on Global Change* (Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, TN, 2002), http://cdiac.esd.ornl.gov/trends/landuse/houghton/houghton.html.
3. C. Prentice *et al.*, in *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, J. T. Houghton *et al.*, Eds. (Cambridge Univ. Press, New York, 2001), pp. 183–237.
4. D. W. R. Wallace, in *Ocean Circulation and Climate*, G. Siedler, J. Church, W. J. Gould, Eds. (Academic Press, San Diego, CA, 2001), pp. 489–521.
5. R. M. Key *et al.*, in preparation.
6. Bottle data and 1° gridded distributions are available through the GLODAP Web site (http://cdiac.esd.ornl.gov/oceans/glodap/Glodap_home.htm).
7. N. Gruber, J. L. Sarmiento, T. F. Stocker, *Global Biogeochem. Cycles* **10**, 809 (1996).
8. C. L. Sabine *et al.*, *Global Biogeochem. Cycles* **13**, 179 (1999).
9. C. L. Sabine *et al.*, *Global Biogeochem. Cycles* **16**, 1083, 10.1029/2001GB001639 (2002).
10. K. Lee *et al.*, *Global Biogeochem. Cycles* **17**, 1116, 10.1029/2003GB002067 (2003).
11. Materials and methods are available as supporting material on *Science* Online.
12. R. Revelle, H. E. Suess *Tellus* **9**, 18 (1957).
13. T. Takahashi, J. Olafsson, J. G. Goddard, D. W. Chipman, S. C. Sutherland, *Global Biogeochem. Cycles* **7**, 843 (1993).
14. Five intermediate water masses and NADW were defined using the following temperature (*T*) and salinity (*S*) properties: Pacific AAIW: 33.8 < *S* < 34.5 and 2 < *T* < 10; NPIW: *S* < 34.3 and 5 < *T* < 12; Indian AAIW: 33.8 < *S* < 34.5 and 2 < *T* < 10; Red Sea Water: *S* > 34.8 and 5 < *T* < 14; Atlantic AAIW: 33.8 < *S* < 34.8 and 2 < *T* < 6; NADW: 34.8 < *S* < 35 and 1.5 < *T* < 4. Water mass inventories were determined by summing up the gridded anthropogenic CO_2 values within a region defined by the *T* and *S* limits using the Levitus World Ocean Atlas 2001 salinity and temperature fields.
15. A. Papaud, A. Poisson, *J. Mar. Res.* **44**, 385 (1986).
16. S. Mecking, M. Warner, *J. Geophys. Res.* **104**, 11087 (1999).
17. A. Poisson, C.-T. A. Chen, *Deep-Sea Res. Part A* **34**, 1255 (1987).
18. M. Stuiver, P. D. Quay, H. G. Ostlund, *Science* **219**, 849 (1983).
19. G. Marland, T. A. Boden, R. J. Andres, in *Trends: A Compendium of Data on Global Change* (Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, TN, 2003), http://cdiac.esd.ornl.gov/trends/emis/meth_reg.htm.
20. D. M. Etheridge *et al.*, *J. Geophys. Res.* **101**, 4115 (1996).
21. C. D. Keeling, T. P. Whorf, in *Trends: A Compendium of Data on Global Change* (Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, TN, 2004), http://cdiac.esd.ornl.gov/trends/co2/sio-keel.htm.
22. R. S. de Fries, C. B. Field, I. Fung, G. J. Collatz, L. Bounoua, *Global Biogeochem. Cycles* **13**, 803 (1999).
23. C. L. Sabine *et al.*, in *The Global Carbon Cycle: Integrating Humans, Climate, and the Natural World. SCOPE 62*, C. B. Field, M. R. Raupach, Eds. (Island Press, Washington, DC, 2004), pp. 17–44.
24. D. E. Archer, H. Kheshgi, E. Maier-Reimer, *Global Biogeochem. Cycles* **12**, 259 (1998).
25. N. Gruber *et al.*, in *The Global Carbon Cycle: Integrating Humans, Climate, and the Natural World. SCOPE 62*, C. B. Field, M. R. Raupach, Eds. (Island Press, Washington, DC, 2004), pp. 45–76.
26. R. A. Feely *et al.*, *Science* **305**, 362 (2004).
27. We thank all individuals who contributed to the global data set compiled for this project, including those responsible for the hydrographic, nutrient, oxygen, carbon, and chlorofluorocarbon measurements, and the chief scientists. The amount of work that went into collecting, finalizing, and synthesizing these data in a manner that makes a publication like this possible can never be properly acknowledged. This work was funded by grants from NOAA/U.S. Department of Energy and NSF. Partial support for K.L. was also provided by the Advanced Environmental Biotechnology Research Center at Pohang University of Science and Technology. This is Pacific Marine Environmental Laboratory contribution number 2683.

Mapping the Antigenic and Genetic Evolution of Influenza Virus

Derek J. Smith,[1,2]*† Alan S. Lapedes,[3]* Jan C. de Jong,[2] Theo M. Bestebroer,[2] Guus F. Rimmelzwaan,[2] Albert D. M. E. Osterhaus,[2] Ron A. M. Fouchier[2]*

The antigenic evolution of influenza A (H3N2) virus was quantified and visualized from its introduction into humans in 1968 to 2003. Although there was remarkable correspondence between antigenic and genetic evolution, significant differences were observed: Antigenic evolution was more punctuated than genetic evolution, and genetic change sometimes had a disproportionately large antigenic effect. The method readily allows monitoring of antigenic differences among vaccine and circulating strains and thus estimation of the effects of vaccination. Further, this approach offers a route to predicting the relative success of emerging strains, which could be achieved by quantifying the combined effects of population level immune escape and viral fitness on strain evolution.

Much of the burden of infectious disease today is caused by antigenically variable pathogens that can escape from immunity induced by prior infection or vaccination. The degree to which immunity induced by one strain is effective against another is mostly dependent on the antigenic difference between the strains; thus, the analysis of antigenic differences is critical for surveillance and vaccine strain selection. These differences are measured in the laboratory in various binding assays (1–3). Such assays give an approximation of antigenic differences, but are generally considered unsuitable for quantitative analyses. We present a method, based on the fundamental ideas described by Lapedes and Farber (4), that enables a reliable quantitative interpretation of binding assay data, increases the resolution at which antigenic differences can be determined, and facilitates visualization and interpretation of antigenic data. We used this method to study quantitatively the antigenic evolution of influenza A (H3N2) virus, revealing both similarities to, and important differences from, its genetic evolution.

Influenza viruses are classic examples of antigenically variable pathogens and have a seemingly endless capacity to evade the immune response. Influenza epidemics in humans cause an estimated 500,000 deaths worldwide per year (5). Antibodies against the viral surface glycoprotein hemagglutinin (HA) provide protective immunity to influenza virus infection, and this protein is therefore the primary component of influenza vaccines. However, the antigenic structure of HA has changed significantly over time, a process known as antigenic drift (6), and in most years, the influenza vaccine has to be up-

---

[1]Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK. [2]National Influenza Center and Department of Virology, Erasmus Medical Center, Dr. Molewaterplein 50, 3015GE Rotterdam, Netherlands. [3]Theoretical Division, T-13, MS B213, Los Alamos National Laboratory, Los Alamos, NM 87545, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: dsmith@zoo.cam.ac.uk

dated to ensure sufficient efficacy against newly emerging variants (7, 8). The World Health Organization coordinates a global influenza surveillance network, currently consisting of 112 national influenza centers and four collaborating centers for reference and research. This network routinely characterizes the antigenic properties of influenza viruses using a hemagglutination inhibition (HI) assay (1). The HI assay is a binding assay based on the ability of influenza viruses to agglutinate red blood cells and the ability of animal antisera raised against the same or related strains to block this agglutination (9). Additional surveillance information is provided by sequencing the immunogenic HA1 domain of the HA gene for a subset of these strains. The combined antigenic, epidemiological, and genetic data are used to select strains for use in the vaccine.

Retrospective quantitative analyses of the genetic data have revealed important insights into the evolution of influenza viruses (10–13). However, the antigenic data are largely unexplored quantitatively because of difficulties in interpretation, even though antigenicity is a primary criterion for vaccine strain selection and is thought to be the main driving force of influenza virus evolution. When antigenic data have been analyzed quantitatively, it has usually been with the methods of, or methods equivalent to, numerical taxonomy (14–16). These methods have provided insights (15–19); however, they sometimes give inconsistent results, do not properly interpret data that are below the sensitivity threshold of the assay, and approximate antigenic distances between strains in an indirect way [discussed by (4, 16, 18)]. Lapedes and Farber (4) solved these problems with a geometric interpretation of binding assay data, in which each antigen and antiserum is assigned a point in an "antigenic map" [based on the theoretical concept of "shape space" (20–23)], such that the distance between an antigen and antiserum in the map directly corresponds to the HI measurement. Lapedes and Farber used ordinal multidimensional scaling (MDS) (24) to position the antigens and antisera in the map.

The method used in this manuscript is based on the fundamental ideas described by Lapedes and Farber (4) and, in particular, takes advantage of their observation that antigenic distance is linearly related to the logarithm of the HI measurement. Exploiting this observation allowed us to create a new method that is parametric yet still handles HI measurements that are beyond the sensitivity of the HI assay (9). We use a modification of metric MDS (25) to position the antigens and antisera in the map (9). This new approach offers computational advantages over the ordinal approach, including reduced running time and fewer local minima, making it tractable to run on datasets the size of the one used in this manuscript, and on larger datasets.

Antigenic map of human influenza A (H3N2) virus. We applied this method to mapping the antigenic evolution of human influenza A (H3N2) viruses, which became widespread in humans during the 1968 Hong Kong influenza pandemic and have been a major cause of influenza epidemics ever since. Antigenic data from 35 years of influenza surveillance between 1968 and 2003 were combined into a single dataset. We sequenced the HA1 domain of a subset of these virus isolates (26, 27) and restricted the antigenic analysis to these sequenced isolates to facilitate a direct comparison of antigenic and genetic evolution. The resulting antigenic dataset consisted of a table of 79 postinfection ferret antisera by 273 viral isolates, with 4215 individual HI measurements as entries in the table. Ninety-four of the isolates were from epidemics in the Netherlands, and 179 were from elsewhere in the world.

We constructed an antigenic map from this dataset to determine the antigenic evolution of influenza A (H3N2) virus from 1968 to 2003 (Fig. 1). Because antigen-antiserum distances in the map correspond to HI values, it was possible to predict HI values that were missing in the original dataset and subsequently to measure those values using the HI assay, so as to determine the resolution of the map. We predicted and then measured 481 such HI values with an average absolute prediction error of 0.83 (SD 0.67) units (each unit of antigenic distance corresponds to a twofold dilution of antiserum in the HI assay) and a correlation between predicted and measured values of 0.80 ($p \ll 0.01$). The accuracy of these predictions indicates that the map has resolution higher than that previously considered available from HI data and higher than the resolution of the assay. The resolution of the map can be greater than the resolution of the assay because the location of a point in the map is fixed by measurements to multiple other points, thereby averaging out errors (9).

The map reveals high-level features of the antigenic evolution of influenza A (H3N2) virus. The strains tend to group in clusters rather than to form a continuous antigenic lineage, and the order of clusters in the map is mostly chronological; from the original Hong Kong 1968 (HK68) cluster, to the most recent Fujian 2002 (FU02) cluster. The antigenic distance from the HK68 cluster, through consecutive cluster centers, to the FU02 cluster is 44.6 units, and the average antigenic distance between the centers of consecutive clusters is 4.5 (SD 1.3) units. The influenza vaccine is updated between influenza seasons when there is an antigenic difference of at least 2 units between the vaccine strain and the strains expected to circulate in the next season; thus, not unexpectedly, we find at least one vaccine strain in each cluster.

The ability to define antigenic clusters allows us to identify the amino acid substitutions that characterize the difference between clusters (Table 1, fig. S1). Some of these "cluster-difference" substitutions (9) will contribute to the antigenic difference between clusters, some may be compensatory muta-
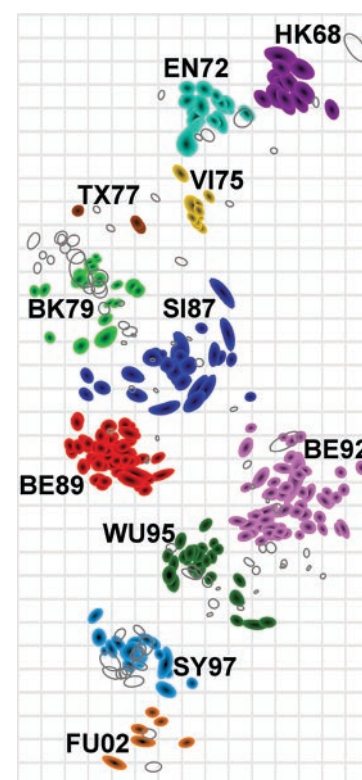

Fig. 1. Antigenic map of influenza A (H3N2) virus from 1968 to 2003. The relative positions of strains (colored shapes) and antisera (uncolored open shapes) were adjusted such that the distances between strains and antisera in the map represent the corresponding HI measurements with the least error (9). The periphery of each shape denotes a 0.5-unit increase in the total error; thus, size and shape represent a confidence area in the placement of the strain or antiserum. Strain color represents the antigenic cluster to which the strain belongs. Clusters were identified by a k-means clustering algorithm (9) and named after the first vaccine-strain in the cluster—two letters refer to the location of isolation (Hong Kong, England, Victoria, Texas, Bangkok, Sichuan, Beijing, Wuhan, Sydney, and Fujian) and two digits refer to year of isolation. The vertical and horizontal axes both represent antigenic distance, and, because only the relative positions of antigens and antisera can be determined, the orientation of the map within these axes is free. The spacing between grid lines is 1 unit of antigenic distance—corresponding to a twofold dilution of antiserum in the HI assay. Two units correspond to fourfold dilution, three units to eightfold dilution, and so on.

tions to retain function, and others may be hitchhikers carried along by chance. Of the 67 cluster-difference amino acid substitutions, 63 were in antigenic sites (28), 8 were in the receptor-binding site (29), and 21 were in codons previously identified as positively selected in an independent genetic dataset covering 1985 to 1997 (10). We see two patterns with respect to these positively selected codons: For the cluster transitions that happened during the period from 1985 to 1997—the period of the sample used to calculate the positively selected codons—most (10 of 12) of the cluster-difference substitutions were in positively selected codons; whereas outside of this timeframe, most [44 of 55, or 16 of 20 if the underrepresented Texas 1977 (TX77) and FU02 clusters are excluded] were not in positively selected codons. A possible explanation for this difference is that cluster-difference substitutions are positively selected, but that the positively selected codons have changed over time, resulting in some pre-1985 positively selected codons not being previously identified, possibly because they were underrepresented in the dataset

used by Bush *et al.* (10). Other possible explanations are that not all cluster-difference substitutions are positively selected or that they cannot be detected as such with methods that use only genetic information.

Comparison of antigenic and genetic evolution. To further investigate the genetic basis of the antigenic cluster structure, we generated a maximum likelihood (ML) phylogenetic tree and a "genetic map" of the HA1 sequences of strains used in the antigenic analysis and color-coded both according to the clusters identified in the antigenic map of Fig. 1 (Fig. 2). The genetic map facilitates a side-by-side comparison with the antigenic map and is a visualization of the amino acid distance matrix calculated from the alignment of HA1 sequences (9). Previous comparisons of the antigenic and genetic evolution have revealed important insights (6, 19, 30, 31); however, a quantitative comparison has not been possible until now because of the previously low resolution of the antigenic data.

We find a remarkable overall correspondence between the relative positions of clusters in the genetic and antigenic maps (Fig. 2,

B and C, respectively). The correlation between antigenic distance and the number of amino acid substitutions between strains was 0.81, and on average, 2.9 amino acid substitutions resulted in one unit change in antigenic distance. The rate of antigenic evolution per amino acid substitution was slower within clusters [on average 3.1 (SD 0.06) amino acid substitutions for each unit of antigenic change] than between clusters [on average 2.1 (SD 0.17) amino acid substitutions for each unit of antigenic change].

There is also a correspondence between the phylogenetic tree and antigenic map, with closely related nucleotide sequences generally belonging to the same antigenic cluster (Fig. 2, A and C). The correlation between antigenic distance and ML phylogenetic tree distance between strains was 0.78, and on average, an ML distance of 0.0085 corresponded to a 1-unit change in antigenic distance.

Although antigenic clusters are mostly contiguous when shown on the phylogenetic tree and genetic map (Fig. 2, A and B), it is not possible to reliably determine antigenic clusters from these genetic data alone. From the tree, it

Table 1. Cluster-difference amino acid substitutions, and distances between antigenic clusters. Cluster-difference amino acid substitutions defined in (9), antibody binding sites defined by (28). Substitution *at a codon in the receptor binding site (29); †at a codon with a rapid rate of amino acid replacement but not positively selected (11); and ‡at a positively selected codon (10). The TX77 and FU02 clusters are represented by fewer strains than other clusters; thus, the number of cluster-difference substitutions into and out of these clusters might decrease with more strains in these clusters. Cluster transitions follow the chronological order of cluster dominance, which is occasionally different from the genetic lineage. Antigenic and genetic distances are between cluster centroids in the antigenic map (Fig. 1) and genetic map (Fig. 2B), respectively. The average standard error (SE) for the genetic distances between clusters was 0.9, for the antigenic distances between clusters was 0.3, and for the ratio was 0.3 (table S1).

Cluster transition	Genetic distance (aa changes)	Antigenic distance (units)	Genetic antigenic ratio	Cluster-difference substitutions					
				Site A	Site B	Site C	Site D	Site E	Other
HK68-EN72	12.1	3.4	3.6	T122N G144D	T155Y* N188D		R207K		
EN72-VI75	14.6	4.4	3.3	N137S*† S145N‡	L164Q Q189K S193D‡	N53D I278S	F174S R102K‡ I213V I217V I230V		
V175-TX77	14.8	3.4	4.4	S137Y*†	G158E‡ Q164L D193N‡	K50R† D53N	S174F K201R‡ V213I V230I	E82K M260I	
TX77-BA79	16.0	3.3	4.8	N133S‡ P143S G146S	K156E‡ T160K Q197R‡	N53D N54S	D172G† V217I V244L	162K K82E	
BA79-SI87	11.9	4.9	2.4	G124D‡	Y155H* K189R				
SI87-BE89	6.9	4.6	1.5	N145K‡					
BE89-BE92	13.7	7.8	1.8	S133D‡ K145N‡	E156K‡ E190D*‡			T262N‡	
BE92-WU95	9.9	4.6	2.2	N145K‡					
WU95-SY97	16.0	4.7	3.4		K156Q‡ E158K‡ V196A†	N276K†		K62E	
SY97-FU02	16.0	3.5	4.5	A131T	H155T* Q156H‡	R50G†		H75Q E83K	L25I V202I W222R G225D*
Total	131.9	44.6							
Average	13.2	4.5	3.2						
SD	2.9	1.3	1.1						

is rarely obvious if a branch or lineage belongs to the same or a different antigenic cluster as its neighbors, and from the genetic map, it is not always possible to determine where one antigenic cluster ends and another begins. The most striking example is the distance between the Sichuan 1987 (SI87) and Beijing 1989 (BE89) clusters, which are genetically closely related but antigenically distinct. The difficulties with an antigenic interpretation of genetic data include the variation in the antigenic effect of amino acid substitutions because of the particular amino acid substitution, the location of the substitution, or the interaction of multiple substitutions.

Surprisingly, a single amino acid substitution, N145K (32), is the only cluster-difference substitution between the SI87 and BE89 and between the Beijing 1992 (BE92) and Wuhan 1995 (WU95) clusters. This is surprising because other cluster transitions are characterized by multiple cluster-difference substitutions and because, on average, a single amino acid substitution causes only 0.37 units of antigenic change. Three pieces of evidence, however, indicate that N145K has a large antigenic effect and, thus, alone can be responsible for a cluster transition. First, there are 12 pairs of strains in the dataset that only differ by N145K, and the average antigenic distance between these pairs in the antigenic map is 4.0 units (SD 1.1). In contrast, other amino acid substitutions at the same position (I145S, N145S), and the same substitution at a different position (N92K), each resulted in less than 1 unit of antigenic change. Second, we took a strain from the BE92 cluster and performed experimental site-directed mutagenesis of position 145 from N to K, and this resulted in 2.6 units of antigenic difference. Third, there were nine strains in the genetic map for which the genetic cluster did not correspond with the antigenic cluster, and N145K was responsible for the difference. These nine strains were interdigitated between the BE92 and WU95 clusters: Five strains from the BE92 antigenic cluster were genetically WU95-like but lacked the N145K substitution [seen as pink triangles in the green WU95 genetic cluster (Fig. 3)], and vice versa, four strains from the WU95 antigenic cluster were genetically BE92-like but had the N145K substitution [shown as green circles in the pink BE92 genetic cluster (Fig. 3)]. To exclude the possibility of laboratory errors, we resequenced and regenerated the HI data for seven of these interdigitated strains and obtained the same results. These three pieces of evidence indicate that a single amino acid substitution, in this case N145K, can cause sufficient antigenic change to be responsible for a cluster transition. Thus, although there is a remarkable correspondence between the genetic and antigenic evolution, there are exceptions that have epidemiological significance of sufficient magnitude that they require an update of the vaccine strain.

Gradual genetic evolution, but punctuated antigenic evolution. A season-by-season analysis of the clusters in the antigenic map shows that in some seasons strains were isolated from more than one antigenic cluster (Fig. 4A). On average, clusters remained dominant for 3.3 years (SD 1.9), with two clusters being dominant for only one season and one for eight seasons. In this dataset, we see strains appear in a cluster up to 2 years before, and 2 years after, the period in which that cluster is the dominant cluster.

The corresponding season-by-season analysis of ML tree distances (Fig. 4B) shows that the rate of genetic change is relatively continuous compared with the rate of antigenic change (Fig. 4A), which is more punctuated. Because this relatively continuous rate of change may in part be due to silent nucleotide substitutions, we repeated the analysis using the number of amino acid substitutions between strains instead of the ML tree distance (Fig. 4C) and found gaps between some clusters, but still a gradual accumulation of mutations, which is not reflected in the corresponding antigenic figure. This finding suggests that some of these amino acid substitutions have little antigenic effect or an effect spreading the cluster sideways in relation to the distance from A/Bilthoven/16190/68 antigen.

The average rates of evolution are given by the slope of the best linear fit to the data in Fig. 4, A, B, and C. The average rate of antigenic drift calculated this way was 1.2 units per year, the average rate of amino acid substitutions was 3.6 per year, and the average rate of change in ML distance was 0.0060 per year. Sometimes the rate of antigenic evolution was faster than genetic evolution and sometimes vice versa, as shown by the deviations from the linear regression line in
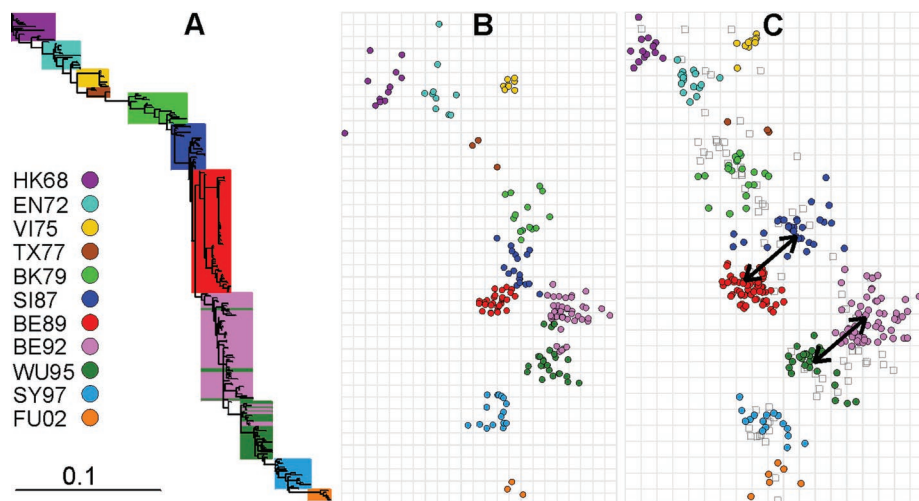
Fig. 2. Comparison of antigenic and genetic evolution of influenza A virus. (**A**) Phylogenetic tree of the HA1 nucleotide sequences, color-coded based on antigenic clusters of Fig. 1. Multiple trees were built using a reversible site-dependent nucleotide ML method (37). There was good consensus among trees, and the tree with ML is shown. (**B**) Genetic map of the HA1 amino acid sequences, color-coded according to the antigenic clusters of Fig. 1. The vertical and horizontal axes represent genetic distance, in this case the number of amino acid substitutions between strains; the spacing between grid lines is 2.5–amino acid substitutions. The orientation of the map was chosen to match the orientation of the antigenic map in Fig. 1. (**C**) The same antigenic map of influenza A virus strains as shown in Fig. 1, except for a rigid-body rotation and translation of the pre-TX77 clusters (fig. S2) to match the genetic map and except that virus strains are represented by colored circles and antisera by open squares. Arrows indicate the two cluster transitions for which the amino acid substitution N145K is the only cluster-difference substitution (Table 1, fig. S1).
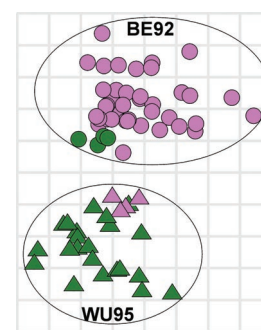

Fig. 3. Detail of the genetic map (Fig. 2B) showing the BE92 and WU95 clusters and how a single amino acid substitution can determine the antigenic cluster. Pink and green symbols represent strains from the BE92 and WU95 antigenic clusters, respectively. Ovals are drawn around the BE92 (circles) and WU95 (triangles) genetic clusters. Green symbols have a lysine (K) at position 145, whereas pink symbols have an asparagine (N) at 145. This single N145K substitution can cause an antigenic cluster change and warrant an update of the vaccine. Two pink triangles are coincident thus only four of the five can be seen. Grid and axes are the same as for Fig. 2B.

Fig. 4D, again, this indicates a remarkable correspondence, with significant exceptions.

The observed pattern of clustered antigenic drift with similar antigenic distances between consecutive cluster centroids is similar to that observed by Gog and Grenfell in a theoretical model in which strain dynamics were governed by a combination of epidemiology and cross-reactive immunity based on antigenic distance (33). This similarity is phenotypic evidence that escape from immunity in the human population plays a major role in determining influenza strain dynamics. Furthermore, there is a selective advantage for clusters that move away linearly from previous clusters as they most effectively escape existing population-level immunity, and this is a plausible explanation for the somewhat linear antigenic evolution in regions of the antigenic map (Fig. 1). The observed deviations from a linear path, as well as the rate of the antigenic evolution, might be determined by tradeoffs between intrinsic viral fitness and extrinsic fitness determined by population-level immunity, possibly in concert with stochastic seeding processes (34), short-lived broad immunity (13), and the phylodynamics of the virus (35).

Genetic analyses of Darwinian selection on influenza HA have focused on the gene level, with more recent refinements to the codon level (10). The quantification of antigenic data described here, which allows the estimation of the antigenic effect of individual amino acid substitutions, provides the opportunity for analyses which integrate selection at the phenotypic level with genetic change at the level of individual amino acid substitutions.

In summary, we have presented a quantification and visualization of the antigenic evolution of influenza A (H3N2) virus from 1968 to 2003 and have tested the accuracy of the method using blind prediction. We show that antigenic evolution is clustered and mostly two-dimensional and reveal a higher rate of antigenic evolution between clusters than within clusters, a remarkable correspondence between antigenic and genetic evolution, but with important exceptions of epidemiological significance, and punctuated antigenic evolution compared with more continuous genetic evolution. The data used for this study were collected as part of routine influenza surveillance, and although there are significant biases in such data, these biases do not have a significant effect on the results (9). This is the most detailed characterization of a real antigenic shape space to date.

From a public health perspective, these methods increase the value of surveillance data and facilitate vaccine strain selection by allowing a finer-grain interpretation of antigenic data, a way to interpret complex data in a simple visual format, and a further integra-tion of antigenic and genetic data. In addition, antigenic maps, in conjunction with strain prevalence data, could be used to quantify the extent to which emerging strains escape immunity in the human population. This would allow the immune-escape component of viral fitness to be compared among multiple cocirculating strains and, if immune-escape is a dominant aspect of total fitness, to be a predictor of which strains would be more likely to seed a new epidemic. It might also be possible to increase the efficacy of repeated vaccination by accounting quantitatively for the antigenic distances among vaccine and circulating strains (23).

We have used human influenza A (H3N2) virus to develop and validate analyses of antigenic properties from binding assay data. We have applied the same methods to the characterization of human H1N1, swine H3N2, and equine H3N8 influenza A viruses, as well as human influenza B virus. There are no assumptions that limit the use of these methods to influenza virus and the HI assay, and we have also applied the methods to the recognition of epitopes by cytotoxic T lymphocytes. We expect these methods will apply to other binding assays such as virus neutralization, complement fixation, and ELISA (36) and that they will be useful for a wide variety of antigenically variable pathogens including human immunodeficiency virus and hepatitis C virus. In general, these methods facilitate the analyses of phenotypes similarly to the way phylogenetic algorithms facilitate analyses of genotypes. Such quantitative analyses have potentially wide-ranging implications for strain surveillance and vaccine strain selection, and for applied and basic research involving antigenically variable pathogens.
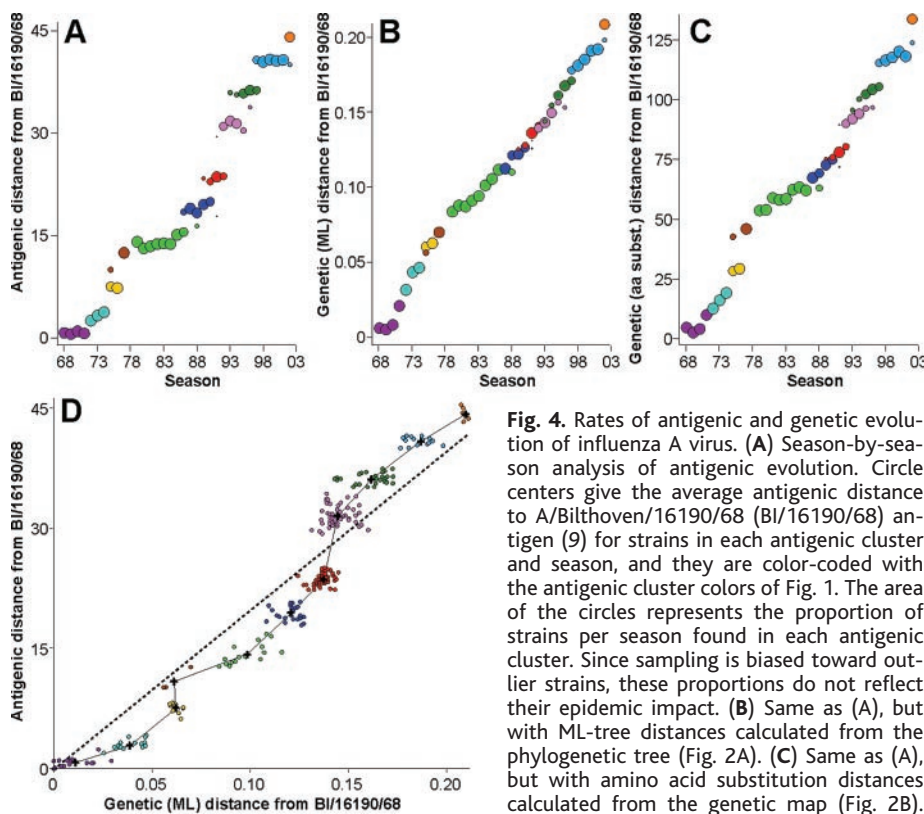

Fig. 4. Rates of antigenic and genetic evolution of influenza A virus. (**A**) Season-by-season analysis of antigenic evolution. Circle centers give the average antigenic distance to A/Bilthoven/16190/68 (BI/16190/68) antigen (9) for strains in each antigenic cluster and season, and they are color-coded with the antigenic cluster colors of Fig. 1. The area of the circles represents the proportion of strains per season found in each antigenic cluster. Since sampling is biased toward outlier strains, these proportions do not reflect their epidemic impact. (**B**) Same as (A), but with ML-tree distances calculated from the phylogenetic tree (Fig. 2A). (**C**) Same as (A), but with amino acid substitution distances calculated from the genetic map (Fig. 2B). (**D**) Comparison of antigenic and genetic evolution. Points are color-coded according to antigenic cluster. The solid line connects the cluster centroids, the dashed line is the best linear fit to the data with a forced zero-intercept.

References and Notes
1. G. K. Hirst, *J. Exp. Med.* **78**, 407 (1943).
2. D. D. Richman, T. Wrin, S. J. Little, C. J. Petropoulos, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 4144 (2003).
3. D. H. O'Connor et al., *Nature Med.* **8**, 493 (2002).
4. A. Lapedes, R. Farber, *J. Theor. Biol.* **212**, 57 (2001).
5. K. Stohr, *Lancet Infect. Dis.* **2**, 517 (2002).
6. A. W. Hampson, in *Influenza*, C. W. Potter, Ed. (Elsevier, London, 2002), pp. 49–85.
7. J. E. Salk, P. C. Suriano, *Am. J. Public Health* **39**, 345 (1949).
8. E. D. Kilbourne et al., *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10748 (2002).
9. Materials and methods are available as supporting material on *Science* Online. For software, see www.antigenic-cartography.org.
10. R. M. Bush, W. M. Fitch, C. A. Bender, N. J. Cox, *Mol. Biol. Evol.* **16**, 1457 (1999).
11. R. M. Bush, C. A. Bender, K. Subbarao, N. J. Cox, W. M. Fitch, *Science* **286**, 1921 (1999).
12. J. B. Plotkin, J. Dushoff, S. A. Levin, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6263 (2002).
13. N. M. Ferguson, A. P. Galvani, R. M. Bush, *Nature* **422**, 428 (2003).
14. P. H. A. Sneath, R. R. Sokal, *Numerical Taxonomy* (Freeman, San Francisco, 1973).
15. T. F. Weijers et al., *J. Virol. Methods* **10**, 241 (1985).
16. D. J. Alexander et al., *Avian Pathol.* **26**, 399 (1997).
17. P. A. Underwood, *J. Gen. Virol.* **62** (Pt. 1), 153 (1982).

18. A. Dekker, G. Wensvoort, C. Terpstra, *Vet. Microbiol.*
47, 317 (1995).
19. J. M. Daly *et al.*, *J. Gen. Virol.* **77** (Pt. 4), 661 (1996).
20. A. S. Perelson, G. F. Oster, *J. Theor. Biol.* **81**, 645 (1979).
21. L. Edelstein, R. Rosen, *J. Theor. Biol.* **73**, 181 (1978).
22. D. J. Smith, S. Forrest, A. S. Perelson, in *Artificial Immune Systems and Their Applications*, D. Dasgupta, Ed. (Springer Verlag, Berlin, 1998), pp. 105–114.
23. D. J. Smith, S. Forrest, D. H. Ackley, A. S. Perelson, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14001 (1999).
24. J. Kruskal, *Psychometrika* **29**, 115 (1964).
25. R. Shepherd, *Hum. Factors* **5**, 33 (1963).
26. Newly sequenced HA1 domain accession numbers: AY660991-AY661211. Previously published HA1 domain accession numbers are in the supporting material.
27. C. A. Macken, H. Lu, J. Goodman, L. Boykin, in *Options for the Control of Influenza IV*, A. D. M. E. Osterhaus, N. Cox, A. W. Hampson, Eds. (Elsevier Science B.V., Amsterdam, 2001), pp. 103–106.
28. D. C. Wiley, I. A. Wilson, J. J. Skehel, *Nature* **289**, 373 (1981).
29. I. A. Wilson, J. J. Skehel, D. C. Wiley, *Nature* **289**, 366 (1981).
30. G. W. Both, M. J. Sleigh, N. J. Cox, A. P. Kendal, *J. Virol.* **48**, 52 (1983).
31. A. J. Hay, V. Gregory, A. R. Douglas, Y. P. Lin, *Philos. Trans. R. Soc. London B. Biol. Sci.* **356**, 1861 (2001).
32. Single-letter abbreviations for the amino acid residues are as follows: I, Ile; K, Lys; N, Asn; and S, Ser.
33. J. R. Gog, B. T. Grenfell, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 17209 (2002).
34. J. R. Gog, G. F. Rimmelzwaan, A. D. Osterhaus, B. T. Grenfell, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 11143 (2003).
35. B. T. Grenfell *et al.*, *Science* **303**, 327 (2004).
36. D. A. Lennette, in *Diagnostic Procedures for Viral, Rickettsial and Chlamydial Infections*, E. H. Lennette, D. A. Lennette, E. T. Lennette, Eds. (American Public Health Association, Washington, DC, 1995).
37. B. Korber *et al.*, *Science* **288**, 1789 (2000).
38. We thank R. van Beek, W. Beyer, T. Bhattacharya, R. Bush, N. Cox, B. Grenfell, J. Gog, H. Gutowitz, A. Hay, R. Hightower, T. Jones, A. Tang, the contributors to

REPORTS

Phase-Resolved Spectroscopy of Geminga Shows Rotating Hot Spot(s)

P. A. Caraveo,[1]* A. De Luca,[1] S. Mereghetti,[1] A. Pellizzoni,[1]
G. F. Bignami[2,3,1]

Isolated neutron stars are seen in x-rays through their nonthermal and/or surface thermal emissions. X-ray Multimirror Mission–Newton observations of the Geminga pulsar show a 43–electron volt spectrum from the whole neutron star surface, as well as a power-law component above 2 kiloelectron volts. In addition, we have detected a hot (170 electron volts) thermal emission from an ~60-meter-radius spot on the pulsar's surface. Such a thermal emission, only visible at selected phase intervals, may be coming from polar hot spot(s), long thought to exist as a result of heating from magnetospheric accelerated particles. It may provide the missing link between the x-ray and gamma-ray emission of the pulsar.

Photons emitted by pulsars carry the signature of their production mechanisms as well as of the geometry of their emitting regions. Although neutron star physics is reflected in their photon spectra, geometrical constraints, such as viewing angles of rotational and magnetic axes, shape their observed light curves. Phase modulation takes place as different emitting regions are brought into view during the star rotation. Geometry can also influence source spectral shapes because of different emission mechanisms in different regions.

[1]Istituto di Astrofisica Spaziale e Fisica Cosmica, Consiglio Nazionale delle Ricerche, Via Bassini, 15-20133 Milano, Italy. [2]Centre d'Étude Spatiale des Rayonnements, Centre National de la Recherche Scientifique, Université Paul Sabatier, 9 Avenue du Colonel Roche, Toulouse, France. [3]Università di Pavia, Dipartimento Fisica Teorica e Nucleare, Via Ugo Bassi, 6 Pavia, Italy.

*To whom correspondence should be addressed. E-mail: pat@mi.iasf.cnr.it

In spite of the potential interest of phase-resolved spectroscopy, the paucity of detected x-ray photons has made it impossible to apply this method to isolated neutron stars (INS), with the exception of the Crab pulsar (*1*). The European Photon Imaging Camera (EPIC) on XMM-Newton and Chandra can now provide an adequate harvest of time-tagged photons. However, phase-resolved spectroscopy is not yet commonly used. So far, it has been applied only to the Crab pulsar with Chandra (*2*) and to 1E1207-5209 with EPIC (*3*, *4*). Although interesting, these sources represent specific and somewhat extreme cases among x-ray–emitting neutron stars. The Geminga pulsar, on the other hand, is often considered as archetypal (*5*) for middle-aged (350,000 years old) neutron stars, which emit x-rays mostly, but not solely, owing to their surface thermal emission. INS surface temperatures frequently yield radiation

in the x-ray domain (*6*), but keV photons can also be produced by energetic electrons in their strong magnetic fields. Geminga has the interesting characteristic of showing both thermal (*7*) and nonthermal processes to be at work in the sub-keV to several keV range (*8*, *9*). With a photon number more than doubling all previous statistics, and with a wider (0.15 to 8 keV) spectral range, EPIC now offers the chance of a meaningful phase-resolved spectroscopy for Geminga.

XMM-Newton performed a 100-ksec exposure on 4 April 2002 with its three EPIC cameras. The two metal-oxide semiconductor (MOS) cameras (*10*) operated in "full frame" mode, while the positive-negative (pn) camera (*11*) operated in "small window" mode, ideal for accurate timing of source photons. After removing intervals with high particle background and correcting for dead time, we obtain a net exposure of 55.0 ksec for the pn camera and of 76.9 and 77.4 ksec for MOS1 and MOS2, respectively. The EPIC observation yielded a total of 76,850 photons in the energy range $0.15 < E < 8$ keV, the majority of which (52,850 photons) are due to the pn detector. The MOS images have unveiled two tails of diffuse emission that are trailing Geminga and are well aligned with the source proper motion (*12*). Here, we present the analysis of the pn data, which were processed with the XMM-Newton Science Analysis Software (SAS version 5.4.1).

First, we have fitted the time-averaged, total source data (Fig. 1), using a combination of a black body and a power law. In view of the unsatisfactory result, we added a third component, both in the form of a black body and of a power law. The resulting χ^2 improved significantly, which suggests that Geminga's spectrum indeed requires a three-component model.

Mapping the Antigenic and Genetic Evolution of Influenza Virus

Derek J. Smith, Alan S. Lapedes, Jan C. de Jong, Theo M. Bestebroer,
Guus F. Rimmelzwaan, Albert D. M. E. Osterhaus, Ron A. M. Fouchier
published online 24 June 2004
doi: 10.1126/science.1097211

MATERIALS AND METHODS

Antigenic map construction

Influenza antigenic data. Within the framework of the WHO global influenza surveillance network, national influenza centers collect epidemic influenza virus isolates for antigenic characterization using the hemagglutination inhibition (HI) assay. The HI assay is based on the ability of the influenza virus hemagglutinin protein to cause agglutination of red blood cells (RBCs) and of specific antisera to block this reaction. These antisera are generally raised in ferrets by infection with prototypic strains, and a fixed quantity of each epidemic virus is tested against a panel of such reference antisera. Two-fold serial dilutions of the antisera are used to determine the highest dilution able to block the agglutination of RBCs. This dilution is called the HI titer, or HI value. The titers can range from <10 (lower dilutions are not used because of potential nonspecific inhibition, to limit the number of titrations, and to limit the amount of antiserum used in each test) to >10,240 (higher titrations are generally not used to limit the number of titrations). The higher the HI value the greater the antigenic relationship between the epidemic virus and antiserum.

Error function. We construct an antigenic map by minimizing the error function, $E=\sum_{ij}e(D_{ij},d_{ij})$, as a function of the antigen and antisera coordinates in the map, thereby determining the position of the antigens and antisera in the map. D_{ij} is the target distance between antigen i and antiserum j, derived from the HI measurement, H_{ij}, using the relation determined in (S1), $D_{ij} =b_j-log_2(H_{ij})$, with b_j set to the log_2 of the maximum measurement for antiserum j. d_{ij} is the Euclidean distance between the coordinates of antigen i and antiserum j in the map. $e(D_{ij},d_{ij})=(D_{ij}-d_{ij})^2$ when the HI measurement is a numeric value, and $e(D_{ij},d_{ij})=(D_{ij}-1-d_{ij})^2 g(D_{ij}-1-d_{ij})$ when the HI measurement is a thresholded value (typically <*10*). $g(x)=1/(1+e^{-10x})$ is a squashing function which allows $e(D_{ij},d_{ij})$ to contribute to E only if $d_{ij} < D_{ij}-1$. We minimize the error function using multiple random restarts of the conjugant gradient optimization method (S2).

Dimensionality. When constructing an antigenic map, one must not only determine the coordinates of the antigens and antisera, but also the dimensionality of the map. To determine the dimensionality, one typically constructs antigenic maps in different numbers of dimensions and then uses a variety of techniques to determine the number of dimensions that most suit the data. A stringent technique is the prediction of HI values which were missing from the HI data used to construct the map, and then measuring these predicted values using the HI assay—this was the method used here. Prediction error decreased only slightly as the number of dimensions increased from 2 to 5, and the cluster structure of the 2, 3, 4, and 5 dimensional maps were identical, indicating that distortion due to compression from higher dimensions to two dimensions is acceptably small, and that a 2-dimensional antigenic map is a good characterization of this data set.

Each of the dimensions in the higher dimensional maps represents antigenic distance just as in the lower dimensional maps. Mathematically, space is not limited to the usual three dimensions of which we are accustomed to thinking. The mathematical properties of space, such as the Euclidean distance between points, can be calculated in a unique and standard way for any number of dimensions.

The biochemical process of binding is complex (involving among other things complementarity in shape and charge, hydrogen bonding, and van der Waals interactions), and the dimensionality of the space necessary to describe binding has been estimated to be in the range of five to ten dimensions (*S3*). Thus the evolution of a pathogen such as human influenza A(H3N2), which we have shown to be mostly two-dimensional, can be thought of as a two-dimensional manifold (subspace) in a higher dimensional space. This dimension reduction is likely due to the interplay of many factors such as the ecological, epidemiological, and evolutionary processes which likely govern influenza strain dynamics (*S4–S7*). We expect that higher dimensional antigenic maps will be required for some pathogens, potentially corresponding to the pathogen's phylodynamics (*S7*).

<u>Comparison with previous work</u>. Most previous approaches to analyzing antigenic data have used the techniques of numerical taxonomy. The limitations of numerical taxonomy when applied to binding assay data are discussed in the main text and by Lapedes and Farber (*S1*). In addition, numerical taxonomy methods do not produce a metric space in which distance directly corresponds to individual measurements in the original data—instead distance is an aggregate of multiple measurements; thus, individual HI values cannot be directly predicted.

Lapedes and Farber introduced a fundamentally different approach, using a variant of ordinal multidimensional scaling (MDS) (*S8*, *S9*) to reconstruct a "shape-space" including both antigens and antisera (*S1*). This ordinal approach used an error function that was a function of distances in Euclidean space, d_{ij}, (which is in turn a function of coordinates of points in the space) and the *rank* of individual HI measurements. Although ordinal MDS is a *nonparametric* method, it produces a metric space. In this space, antigens and antisera are both represented and thus the results can be tested by prediction of individual HI values. One advantage of the ordinal approach is that it is not necessary to specify a priori the parametric relationship between HI value and distance, indeed this relationship can be determined using ordinal MDS and it turns out that distance is linearly related to the logarithm of the HI value (*S1*).

We have used this linear relationship as the foundation of a new *parametric* method to generate antigenic maps, and thus the relative positions of antigens and antisera in the map can be determined with a variant of metric (instead of ordinal) MDS (*S10*, *S11*). Metric approaches construct an error function that is a function of the distances in Euclidean space and the *value* of individual HI measurements (instead of the *rank* of the HI measurements used in ordinal approaches).

When an HI measurement is beyond the sensitivity of the assay, it is reported as a thresholded value, typically "<10." To account for such relational information we modified the metric approach to be able to deal with both values and relations, as described in the section on the error function above.

Consensus. The value of the error function for the antigenic map of Fig. 1 is 3523.74. Consensus among the different minimizations was high; the relative positions of pre-TX77 strains was highly repeatable, as was the relative positions of post-TX77 strains; however, the relative position of the pre-TX77 region to the post-TX77 region had more than one stable state (Fig. S2). Distances between pre- and post-TX77 strains were excluded when calculating the correlation between antigenic distance and the number of aa substitutions due to this metastability.

Accuracy. Since antigen-antiserum distances in the map correspond to HI values, it was possible to predict HI values that were missing in the original data set, and subsequently measure those values using the HI assay, to determine the resolution of the map. When we blindly predicted 481 such HI measurements, the average magnitude of the prediction error for the two-dimensional map was 0.83 [SD 0.67, correlation between predicted and measured values 0.80 ($P \ll 0.01$)] units. In addition to blind prediction from a 2-dimensional map, we also blindly predicted from 3, 4, and 5 dimensional maps. Prediction error decreased only slightly as the number of dimensions increased, indicating (along with previous evidence that the cluster structure of the 2, 3, 4, and 5 dimensional maps were the same) that distortion due to compression from higher dimensions to 2 dimensions is acceptably small for this data set.

Rigidity. The rigidity of the coordinates of points in an antigenic map depends not only on the number of HI measurements between the antigens and antisera, but also on how well spread the antigens and antisera are in space. In direct analogy to geographic map making, the coordinates of points can be determined most accurately if they can be well triangulated with other points. For example, if three columns of antisera have similar HI values for each of the antigens they are measured against, then those antisera will likely end up in similar locations in the antigenic map, and although they will provide repeated-measure type information to determine the antigenic distance between the antigens and the three antisera, they will provide little information about the distances among the antigens. In contrast, three antisera that are more widely spread will be able to locate the antigens relative to each other, as well as to the antisera. Thus when selecting reference antisera, it is best to choose antisera which are distinct from each other. Once an antigenic map has been made, the accuracy (by prediction described above), rigidity of individual points (described below), and sensitivity to sample by bootstrapping (described below) can be derived and confidence areas delineated; if required, extra HI values can be measured to increase the resolution and tighten the confidence areas. Given antisera and antigens that are well spread in space, an antigenic map with accuracy similar to the one shown in Fig. 1 can be made from relatively small amounts of data. For example, most of the HI tables used in this study contain data predominantly from a single antigenic cluster, and contain antigens measured against typically 6 to 8 antisera, and almost all of these HI tables produce reasonable antigenic maps even when not combined into a single data set.

To determine the rigidity confidence area for each individual antigen and antiserum, all other antigens and antisera were held fixed, and the target antigen or antiserum was displaced to determine how the error function changed due to the displacement. The periphery of the shapes in Fig. 1 indicates an increase of 0.5 in the error function.

Shading illustrates the rate of error increase for each antigen, from black (no error) to the base color of the antigenic cluster (0.5 error) at the periphery.

Sensitivity to the data set sample. We tested the sensitivity of the antigenic map to the particular antigens, antisera, and individual HI measurements by bootstrap resampling of each of these components. We used the standard bootstrap technique of sampling the existing data set with replacement and generating new data sets of the same size as the original data set typically resulting in resamples with multiple entries of some components and missing entries of other components. Other than the same trans-TX77 metastability discussed above, the maps generated from these new data sets were generally robust to such resampling. We also bootstrap resampled the antigens to generate data sets with the same number of antigens per season (separate resamples with 3 and 10 strains per season). Again, apart from the trans-TX77 metastability, the maps were generally robust to the resampling.

One of the major topographical features of the map in Fig. 1 is the change in direction of the antigenic evolution between the 1980s and the 1990s. The timing of this change is somewhat concomitant with the appearance of the BE89 cluster genetic dead-end. To test if this change was an artifact introduced by the BE89 cluster, we remade the map omitting the BE89 antigens and antisera. The resulting map was almost identical to the original, indicating that the change is not an artifact of the BE89 genetic dead-end.

Combining HI data. When combining HI data, either from tables made at different times in the same laboratory, or tables made in different laboratories, the map made from the combined data might have higher or lower resolution than maps made from the individual tables. The accuracy of the individual and combined maps, and individual antigens and antisera in the maps, can be determined by prediction (either by measuring previously unmeasured values, or by leaving out already measured values), calculating rigidity and bootstrap resampling confidence areas as described above. Unless there are systematic differences between the data sets, adding more data will increase the resolution of the map; if there are systematic differences, then they can be detected using these methods. These methods also provide a way to curate an antigenic database and maintain consistency: as new antigenic data are produced, reference antigens and/or antisera can be included, the positions of which are compared in the new and existing maps.

Variation in red blood cells (RBCs) and antisera. Variation in the reagents used in the HI assay, such as the animal source of red blood cells (RBCs), and animal used to raise the antisera, add potential complexity to HI data.

The data used in this study were collected in a single laboratory, all antisera were raised in ferrets, RBCs up to and including the 1991/92 season were from chickens and thereafter from turkeys. As part of the data validation process for this study, we remeasured, using turkey RBCs, 88 titers of strains from 1968 to 1989 that had previously been measured as part of routine surveillance using chicken RBCs. The average titer difference was 0.01 of a 2-fold dilution, and the average of the absolute values of the titer differences was 0.68 of a 2-fold dilution, indicating that the switch from chicken to turkey RBCs is unlikely to have a significant effect in this data set.

Had the switch from chicken to turkey RBCs had a larger effect, the methods described here would provide techniques to examine the differences quantitatively: a map made

with strains only measured using chicken RBCs could be compared to a map with measurements using turkey RBCs, and strains in common between the maps could be compared for differences in antigenic distance to other strains.

Similarly, maps could be made for the same antigens using antisera from different animals to analytically compare differences among such antisera. Hyperimmune sheep (or goat) antisera for example are less specific than ferret antisera; hence, an antigenic map made only with sheep antisera would be scaled differently from a map made only with ferret antisera, and the difference in scale could be determined by comparing the two maps. Furthermore, this difference in scale could then be used to make antigenic maps from HI tables containing measurements using both ferret and sheep antisera (the scale factor would be applied in the $e(D_{ij}, d_{ij})$ part of the error function, which would take a third parameter to indicate the animal source of antisera). Alternatively, the scale factor could be a free parameter in the error function which would be then be determined as part of the error minimization process. If the difference between hyperimmune sheep or goat antisera were not linear, but varied as a function of the antigenic distance between antigens, then this could also be determined using the same methods, and the $e(D_{ij}, d_{ij})$ term in the error function could be extended with a nonlinear function instead of a simple scale factor. In practice, it is likely that once the relation among ferret and hyperimmune sheep or goat antisera has been determined using these methods, it could then be used transparently in future analyses.

Genetic map construction

Our method for making genetic maps is similar to that for making antigenic maps, except that the target distances are the number of aa substitutions between the antigen aa sequences, and there are no antisera points. To avoid underestimating genetic distances due to sequential mutations at the same location, when a target distance was greater than 30 aa substitutions it was allowed to contribute to the error function only when the map distance was <30 in the same way that a thresholded HI measurement was only allowed to contribute to the error function when the map distance was within threshold. PCOORD (*S12*) is a related method for making genetic maps, but does not directly handle thresholded distances. The correlation between the number of aa substitutions and the corresponding distances between strains in the genetic map was 0.98 indicating that the 2D genetic map is a reasonable representation of the target aa distance matrix.

Clustering

Clusters were identified in the antigenic map by a *k*-means clustering algorithm (*S13*), using average weighting, and *k*=10. Two adjustments were then made. First, the three TX77-like strains were made into a separate cluster. TX77 is a well-known antigenic and genetic type (*S14*), and strains of this type were recorded in our laboratory, but apart from the three strains in the data set, the remaining strains, and HI data for these strains, are currently lost. Second, strains VI/7/87, OK/5/88, and VI/1/88 were assigned to the SI87 cluster—in multiple runs of the *k*-means clusterer these strains appeared in either the SI87 or BE89 clusters with very small error difference between placement in either cluster. The most parsimonious placement of these strains, based on genetic and temporal information is in the SI87 cluster.

When clusters were determined from the genetic map (instead of the antigenic map), and plotted season-by-season, the antigenic evolution was still punctuated and the genetic evolution still gradual.

"Cluster-difference" aa substitutions

An aa substitution X to Y at location L is considered a "cluster-difference" substitution between clusters A and B if all (or all but one) strains in cluster A have aa X at location L, and all (or all but one) strains in cluster B have aa Y at location L. Fig S1 shows an alignment of the aa sequences used in this study, grouped and color-coded based on antigenic cluster, and marked with the cluster-difference aa substitutions.

Evolutionary distance to A/Bilthoven/16190/68 antigen

One way to measure the distance of a strain to the root of the 1968 pandemic strains would be to measure the straight-line distance on the map from the strain of interest to a reference strain chosen to be representative of the 1968 pandemic. An obvious choice for such a reference strains would be the A/HongKong/1/68 (HK/1/68) vaccine strain. However, we choose a different strain, A/Bilthoven/16190/68 (BI/16190/68), because in both the ML tree (Fig. 2A) and a protein parsimony tree (data not shown) BI/16190/68 was closest to the avian outgroup, and also because with only 5 passages in tertiary monkey kidney cells and no passages in eggs, it likely has fewer laboratory adaptations than the HK/1/68 vaccine strain and thus is likely to be a better representative of the early pandemic strains. We also measure distance to BI/16190/68, not in a straight-line, but as an evolutionary distance through the centroids of clusters between the strain of interest and BI/16190/68. This path between cluster centroids is uniquely determined for influenza virus as the clusters are chronologically ordered (for other pathogens this route might not be clear). Note, this implies going via the BE89 cluster when passing from the BE92 cluster back to the SI87 cluster, as this is the order in which the clusters appeared, and thus the order which will influence selection at the phenotypic level and the most appropriate measure here of antigenic distance, rather than following the genetic lineage which would bypass the BE89 genetic dead-end. Similarly, the antigenic path from TX77 to EN72 goes through VI75. Taking this evolutionary path, rather than the straight-line distance, to the root antigen is a map analog of measuring genetic distance from a phylogenetic tree rather than as the number of mutations to the root strain—for short distances these measures are usually similar, but for longer distances the straight-line distance will be shorter than the evolutionary distance. The evolutionary distance to the BI/16190/68 strain in the antigenic and genetic maps was thus calculated as the sum of distances of a strain to the centroid of the previous antigenic cluster, from this centroid through the intermediate centroids to the EN72 centroid, and then to BI/16190/68 antigen in the HK68 cluster.

Biases in the data

There are three biases in the data used for this study. First, the majority of strains selected for sequencing as part of routine surveillance were those with unusual antigenic properties. The remaining strains were the first and last isolate from each season, a representative of each antigenically distinct cluster seen in the Netherlands that season (determined each year as part of routine surveillance by classical reading of HI data), and the vaccine strains. This sample is thus likely to contain a larger proportion of outliers than would be obtained from a random sample; thus, the cluster structure of a map made

from a random sample would likely be more clustered than the map of Fig. 1, and the observation of clustered antigenic evolution would be unlikely to change. The removal of outlier strains would be unlikely to alter the overall structure of the map as the map is robust to bootstrap resampling of the antigens and to the removal of whole clusters such as BE89. For a subset of seasons we generated an antigenic map for all strains collected at the Dutch National Influenza Center—this is closer to a random sample than the sequenced strains; the cluster structure of Fig. 1, and the punctuated antigenic evolution of Fig. 4A persisted. As indicated in the caption of Fig. 4, since sampling is biased towards outlier strains, the area of the circles in Figs. 4 A, B and C, does not reflect epidemic impact. The strains in each antigenic cluster were bootstrap resampled to test the robustness of distances between antigenic clusters in the genetic and antigenic maps shown in Table 1. Distances were not statistically significantly different from the distances in original sample, and SEs were sufficiently small to indicate that distances are robust to the surveillance sampling biases (Table S1).

Second, approximately one third of the strains in this study were isolated in a single country—The Netherlands. This sampling bias is likely to affect the season in which the first strains in a cluster are detected, as new variants are likely to circulate locally before their global emergence. A further effect is that some variants, which only circulate locally before dying out, might not appear in this data set. The addition or absence of such local data, as with the removal of outlier strains, would be unlikely to change the overall structure of the map as both the antigenic and genetic maps are robust to the presence or absence of even major clusters.

Third, mutations can be introduced by laboratory propagation of influenza virus strains in embryonated chicken eggs. However, since 89% of strains used in this study were propagated only in mammalian cell culture, for at most 5 passages, and usually only 1 or 2 passages, this bias is unlikely to have a significant effect on this study (*S15, S16*).

Accession numbers
In addition to the newly sequenced HA1 domains, the following previously published HA1 domain sequences were used in this study: AF008665 AF008697 AF008711 AF008725 AF008755 AF008769 AF008828 AF008867 AF008886 AF008888 AF008903 AF008905 AF092062 AF131997 AF180570 AF180602 AF180643 AF201874 AF368444 AF368446 D21173 D49961 ISDN38157 ISDNCDA001 ISDNENG72 ISDNHK71 ISDNTX77 ISDNVIC75 M16739 U08858 Z46405 Z46408 Z46413 Z46414.

Table S1. Distances between antigenic clusters in the genetic and antigenic maps. Data are related to the data in Table 1 of the main text except that distances and ratios are the average of 1,000 bootstrap resamples of the strains in each antigenic cluster, and the standard error (SE) of the averages are shown. Each resample was of the same number of strains per cluster as in the original sample. Distances and ratios differ slightly, but not statistically significantly, from those in Table 1 because these are averages of 1,000 bootstrap resamples, whereas the data in Table 1 were calculated from the original sample.

Cluster transition	Genetic distance (aa substitutions)		Antigenic distance (units)		Genetic/antigenic ratio	
	Av	SE	Av	SE	Av	SE
HK68-EN72	12.19	0.93	3.43	0.23	3.57	0.26
EN72-VI75	14.71	0.54	4.38	0.24	3.37	0.25
VI75-TX77	14.87	0.98	3.41	0.66	4.48	0.65
TX77-BA79	16.08	2.04	3.34	0.27	4.82	0.56
BA79-SI87	11.99	1.11	4.93	0.31	2.44	0.2
SI87-BE89	6.87	0.5	4.60	0.27	1.50	0.12
BE89-BE92	13.73	0.41	7.77	0.19	1.77	0.07
BE92-WU95	9.89	0.75	4.59	0.21	2.17	0.16
WU95-SY97	16.07	0.86	4.69	0.17	3.43	0.23
SY97-FU02	16.05	0.77	3.52	0.35	4.60	0.44
Average SE		0.89		0.29		0.29
Total	132.43	1.79	44.66	0.97		
Av	13.24	0.18	4.47	0.1	3.21	0.13

Figures appear at the endn of the file

Fig S1. Alignment of aa sequences used in this study, grouped and color-coded according to the antigenic clusters, and with the cluster-difference substitutions boxed. The nine strains interdigitated between the BE92 and WU95 clusters are underlined. Codons in which there were no aa substitutions are excluded. The image can be zoomed to show details.

Fig S2. The antigenic map of influenza A virus strains as shown in Fig. 2C, with the rigid-body rotation and translation of the pre-TX77 clusters shown by the arrow. The positions in the original map (Fig. 1) are shown in light colors, and darker colors indicate the positions after the rotation and translation.

Fig S3. The antigenic map of influenza A virus strains as shown in Fig. 1 except that virus strains are represented by colored circles and antisera by open squares. A line connecting a virus strain an antiserum indicates that the virus strain was used to raise the antiserum.

References and Notes

S1.   A. Lapedes, R. Farber, *J Theor Biol* **212**, 57-69 (Sep 7, 2001).
S2.   B. Flannery, S. Teukolsky, W. Vetterling, *Numerical Recipes in C* (Cambridge Univ. Press, Cambridge, 1988).
S3.   A. S. Perelson, G. F. Oster, *J Theor Biol* **81**, 645-70 (1979).
S4.   N. M. Ferguson, A. P. Galvani, R. M. Bush, *Nature* **422**, 428-33 (2003).
S5.   J. R. Gog, B. T. Grenfell, *Proc Natl Acad Sci U S A* **99**, 17209-14 (2002).
S6.   J. R. Gog, G. F. Rimmelzwaan, A. D. Osterhaus, B. T. Grenfell, *Proc Natl Acad Sci U S A* **100**, 11143-7 (2003).
S7.   B. T. Grenfell *et al.*, *Science* **303**, 327-332 (2004).
S8.   J. Kruskal, *Psychometrika* **29**, 1-27 (1964).
S9.   J. Kruskal, *Psychometrika* **29**, 115-129 (1964).
S10.   R. Shepherd, *Human Factors* **5**, 33-48 (1963).
S11.   R. Shepherd, *J Math Psychol* **1**, 54-87 (1964).
S12.   D. G. Higgins, *Comput Appl Biosci* **8**, 15-22 (1992).
S13.   R. Duda, P. Hart, *Pattern Recognition and Scene Analysis* (Wiley Interscience, 1973).
S14.   G. W. Both, M. J. Sleigh, N. J. Cox, A. P. Kendal, *J Virol* **48**, 52-60 (1983).
S15.   J. C. de Jong *et al.*, *Bull World Health Organ* **66**, 47-55 (1988).
S16.   G. C. Schild, J. S. Oxford, J. C. de Jong, R. G. Webster, *Nature* **303**, 706-9 (1983).