# Antibody landscapes after influenza virus infection or vaccination

J. M. Fonville[1,2,3‡], S. H. Wilks[1,2‡], S. L. James[1,2], A. Fox[4], M. Ventresca[1⌣], M. Aban[5], L. Xue[5], T. C. Jones[1,2], Le N. M. H.[4], Pham Q. T[6], Tran N. D.[6], Y. Wong[7], A. Mosterin[1,2], L. C. Katzelnick[1,2], D. Labonte[8], Le T. T.[6], G. van der Net[3], E. Skepner[1,2], C. A. Russell[2,9], T. D. Kaplan[10], G. F. Rimmelzwaan[3], N. Masurel[3†], J. C. de Jong[3], A. Palache[11], W. E. P. Beyer[3], Le Q. M.[6], Nguyen T. H.[6], H. F. L. Wertheim[4,12], A. C. Hurt[5,13], A. D. M. E. Osterhaus[3], I. G. Barr[5], R. A. M. Fouchier[3], P. W. Horby[4,12], D. J. Smith[1,2,3*]

[1]Center for Pathogen Evolution, Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK

[2]WHO Collaborating Center for Modeling, Evolution, and Control of Emerging Infectious Diseases, Cambridge CB2 3EJ, UK

[3]Department of Viroscience, Erasmus Medical Center, Rotterdam 3015 CE, the Netherlands

[4]Oxford University Clinical Research Unit and Wellcome Trust Major Overseas Programme, Hanoi, Vietnam

[5]WHO Collaborating Centre for Reference and Research on Influenza, VIDRL at the Peter Doherty Institute for Infection and Immunity, Melbourne VIC 3000, Australia

[6]National Institute of Hygiene and Epidemiology, Hanoi, Vietnam

[7]Oxford University Museum of Natural History, Oxford OX1 3PW, UK

[8]Insect Biomechanics Group, Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK

[9]Department of Veterinary Medicine, University of Cambridge, Cambridge CB3 0ES, UK

[10]bobblewire.com, Saint Louis, MO 63112, US

[11]Abbott Laboratories, Weesp 1380 DA, the Netherlands

[12]Nuffield Department of Clinical Medicine, Centre for Tropical Medicine, University of Oxford, Oxford OX3 7BN, UK

[13]Melbourne School of Population and Global Health, University of Melbourne, Parkville VIC 3010, Australia

‡ These authors contributed equally to this work.

*Correspondence to: dsmith@zoo.cam.ac.uk.

† Professor Masurel is deceased

⌣ Current address: School of Industrial Engineering, Purdue University, West Lafayette, IN 47907 USA.

# Contents

# 1 Methods

## 1.1 Datasets

### 1.1.1 Ferret data and the hemagglutination inhibition assay

Sera from 35 influenza-naïve ferrets each infected with a single influenza A/H3N2 virus (viruses isolated between 1989 and 2010) were collected 13-20 (typically 14) days post-infection. These sera were tested with the hemagglutination inhibition (HI) assay against 74 A/H3N2 influenza viruses isolated globally between 1992 and 2011; the full titration table is given in Table S1 (see additional excel files). The HI assay is based on the ability of the influenza virus hemagglutinin protein to cause agglutination of red blood cells (RBCs) and of virus-specific antisera to block this reaction (*3*). The ferret antisera were tested against a fixed quantity of each influenza virus (standardized to 4 HA units/25 $\mu$L). Prior to testing, each ferret antiserum was treated to remove non-specific inhibitors, by combining 1 part antiserum with 4 parts receptor destroying enzyme (RDE) (100 units/25$\mu$L) (Seiken RDE II, Denka Seiken Co., Ltd.) and incubating at 37°C overnight. Five parts 1.6% (w/v) sodium citrate solution (Ajax Finechem Pty Ltd.) were then added to the RDE-treated serum and incubated for 30 minutes in a 56°C water bath to inactivate the RDE. The treatment of antiserum resulted in a ten-fold dilution. RDE-treated antiserum was then adsorbed with 5% (v/v) packed turkey RBC in a 10:1 ratio, mixed by inversion and incubated at room temperature for 30 minutes before centrifugation of the RBC pellet at 2000 rpm for 5 minutes followed by removal of the adsorbed serum supernatant. Serial two-fold dilutions of the treated antisera in CaMg-free Phosphate Buffered Saline (PBS), pH 7.2 (Sigma-Aldrich), were used to determine the highest dilution of antiserum capable of blocking agglutination between test influenza viruses and 1% (v/v) turkey RBCs in a 96 well V-bottom Micro Test plate (Greiner Bio-One). This dilution is called the HI titer, or HI value. The lower detection limit of the titers was 10 (lower dilutions were not possible due to the dilution factor associated with RDE treatment). High HI values indicate a close antigenic relationship between virus and antiserum.

To determine the antibody titer of a human serum, the HI assay was executed as above, but with 50 $\mu$L of the treated human serum of interest instead of the ferret antiserum. Following each HI assay, viruses were retitrated to verify that their HA titer was still at 4-8 HA units/25 $\mu$L. Ferret antisera from the Wuhan 1995 (WU95), Sydney 1997 (SY97), Fujian 2002 (FU02) and Wisconsin 2005 (WI05) antigenic clusters, and their homologous antigens with known end-point titers, were included in each assay as positive controls. Day 0 sera and PBS samples were tested in each assay and served as negative controls.

### 1.1.2 Ha Nam cohort data

Based on provisional serological data from titrations against a single virus, influenza-like illness (ILI) reports and polymerase chain reaction (PCR) tests, we selected a subset of individuals from the Ha Nam Household Cohort Study (*12*), monitoring households in Ha Nam, Viet Nam. This subset included all individuals from six households that had a high sampling compliancy and a household size of five or larger (n=36). Additionally, we included individuals that had a PCR-confirmed infection (n=10), a random subset of participants with serological conversion (n=14, defined as a $\geq$4-fold change in HI titer against a representative strain, see Table S2), and a random subset of control individuals without reported serological change (n=9). In total, 69 individuals were selected, resulting in 324 landscapes (37 of whom have the full 6-year time series from 2007-2012). The 36 individuals from the six selected households were titrated with the HI assay against 38 viruses (isolated between 1993-2011), while the remaining 33 individuals were titrated against a more extensive set of 40 viruses to enable further study of the back-boost (viruses isolated between 1968 and 2011), sufficient serum volume permitting; the maximum dilution measured was 640. If agglutination was still inhibited at this dilution, the entry was listed as having an end-point of $\geq$1280, since 1280 represents the next two-fold dilution that would have been measured under our protocol (applicable to 93 out of 11597 titrations). The full titration table is given in Table S3 (see additional excel files).

Individuals in the cohort were closely monitored for signs of ILI. Upon detection of ILI, a swab was taken for PCR analysis (typically taken the next day, or within two days if reported over a weekend) and subtyping (*12*).

During the cohort study period, two A/H3N2 antigenic clusters circulated: first WI05, and later PE09, as indicated in Table S2. To indicate the viruses that may potentially have caused infection in a year, we highlight all viruses belonging to the cluster that circulated in the year prior to each sample collection date in our figures. For example, in samples collected in 2008 this includes all WI05-type viruses, in 2010 it includes all PE09-type viruses, while in 2009 it includes both WI05 and PE09 type viruses, since both clusters circulated that year.

| Year | Sampling date | Cluster(s) circulating | Inclusion test strain |
|---|---|---|---|
| 2007 | 12/2007 | WI05 | - |
| 2008 | 12/2008 | WI05 | A/Vietnam/EL140/2008 |
| 2009 | 06/2009 | WI05, PE09 | A/Vietnam/TX265/2009 |
| 2010 | 04/2010 | PE09 | A/Vietnam/TX265/2009 |
| 2011 | 07/2011 | PE09 | A/Vietnam/TX265/2009 |
| 2012 | 05/2012 | PE09 | A/Vietnam/TX265/2009 |

Table S2: Cohort year, sampling date, circulating antigenic clusters and representative antigenic strains used for seroconversion testing (WI05: Wisconsin 2005; PE09: Perth 2009).

In our data analyses, we used reference viruses against which all subjects were titrated and defined seroconversion as a $\geq$4-fold increase in HI titer against A/Vietnam015/EL134/2008 in years with WI05 circulation, and against A/Perth/16/2009 for PE09 cluster circulation.

None of the 69 subjects had ever received an influenza vaccination. Only subjects that were 5 years old or older were enrolled into the Ha Nam cohort, but otherwise no age restrictions where imposed. The stem-plot in Table S4 gives the subject ages at the end of the study period (2012):

0 — 7889
1 — 0022223455556779
2 — 0122235677889
3 — 22358
4 — 0223345788999
5 — 000012222235
6 — 38
7 — 13
8 —
9 — 15

Table S4: A stem-and-leaf-plot representing the ages of the individuals of the Ha Nam cohort in 2012. A stem-and-leaf plot is similar to a histogram in that the length of the rows indicates the number in each group, and additionally indicates the data within each group. The row 9 — 15, for example, indicates two individuals: one aged 91, the other 95.

Each subject was informed of the aims, methods, anticipated benefits and potential risks of the study by a native Vietnamese-speaking member of the research team in a face-to-face meeting. Participants were provided with age appropriate information sheets in Vietnamese and asked to sign a consent form witnessed by a member of the study team. All participants provided written informed consent.

### 1.1.3 Vaccination studies

Sera of 102 volunteers pre- and post-vaccination with A/Nanchang/933/95 (study performed in 1997) and sera of 123 volunteers pre- and post-vaccination with A/Sydney/5/97 (study performed in 1998) were collected by CSL Ltd. (Parkville, Australia). Sera were collected between 26 and 33 days after vaccination. The sera were titrated with HI against 70 A/H3N2 influenza viruses, serum volume permitting. For this study, the end-point of the titration was always determined. The full table of titrations is given in Table S5 (see additional excel files).

Both these studies consist of a fairly dichotomous population, with the majority of individuals either in their late teens to early twenties, or in their sixties to early seventies (Table S6 and Table S7).

1 — 888899999999999999
2 — 000000000111122222334449
3 — 0
4 — 8
5 —
6 — 00111222223334444444555556677777899999
7 — 000112222333344445

Table S6: A stem-and-leaf-plot representing the ages of the individuals in the 1997 vaccination study.

1 — 888888889999999999999
2 — 000000011111112223344666677788
3 — 12239
4 — 05679
5 —
6 — 122334445566667778888889999
7 — 000000000011112222223333334444555555

Table S7: A stem-and-leaf-plot representing the ages of the individuals in the 1998 vaccination study.

### 1.1.4 Phylogenetic analysis of used A/H3N2 influenza viruses

A phylogenetic tree of the viruses used in this study is shown in Figure S1. In this figure, the linear evolution of influenza A/H3N2 is clear, as shown by a spindly and narrow phylogenetic tree. The tree shows a gradual and consistent accumulation of genetic changes over time. By colour-coding by antigenic cluster, it can also be seen that the antigenic evolution of influenza A/H3N2 viruses has been progressive and step-wise, with antigenically similar viruses circulating for a few years before strains with related but novel antigenic characteristics replace them. For reference, in the HA1 gene alone, 69 out of 346 amino acid positions were changed from the HK68 A/Bilthoven/16190/68 strain until the PE09 vaccine strain A/Victoria/361/2011 (171 nucleotide mutations).

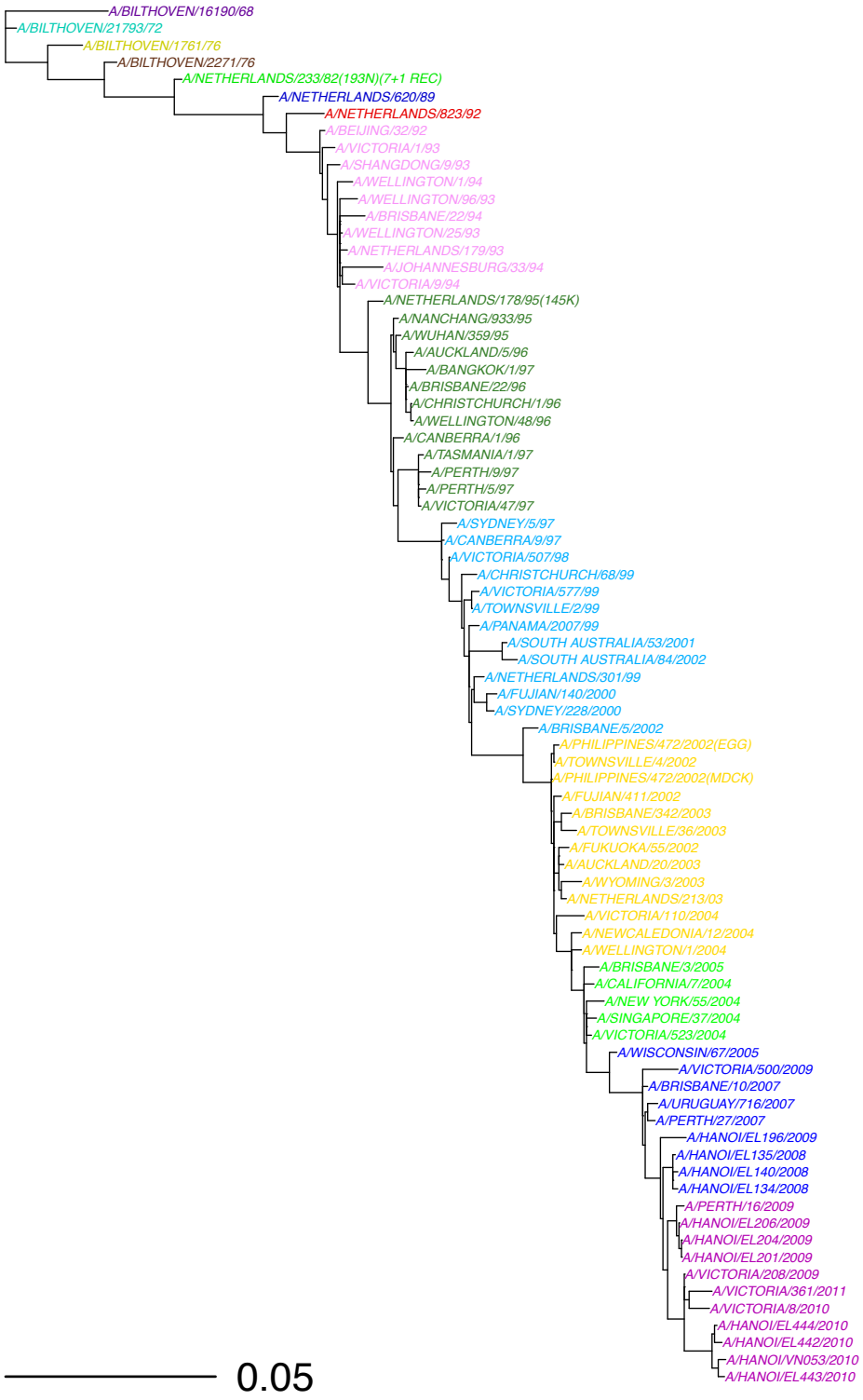



Figure S1: Phylogenetic tree for the nucleotide sequences of the HA gene for the 81 strains used in our studies. The sequences were aligned with MAFFT, and the full HA gene was analysed. The initial tree was constructed with PhyML using the GTR+I+$\Gamma_4$ evolutionary model, and topology and branch length were optimized in GARLI (version 2.0.19) with 5 separate runs each of 20,000 generations. The strain names are colour-coded by antigenic cluster designation.

## 1.2 Creating and testing antibody landscapes

All analyses were performed using R software for statistical computing, unless otherwise specified (version 3.0.3, R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org), the code is available from the authors on request.

### 1.2.1 Determining the antigenic map

Antigenic cartography positions sera and viruses in a map based on their HI titrations such that distance between viruses represents their antigenic relationship (*2*). This methodology, as implemented in the antigenic cartography software (lisp-implementation), was used to calculate the coordinates of the viruses based on the titrations with ferret sera from Table S1. The positions of 74 viruses were found by minimizing the mapping error. An "overlay-merge" was made with the Smith *et al.* antigenic map (*2*), which minimizes the squared error of positioning of viruses in common between the two data sets (9 viruses), to obtain the coordinates of the 7 antigenic cluster prototype viruses that were in the Smith *et al.* antigenic map, but not among the viruses titrated for this study (HK68, EN72, VI75, TX77, BK79, SI87 and BE89). The resulting antigenic map is shown in Figure S2, and the antigenic coordinates computed for each virus can be found in Table S8.

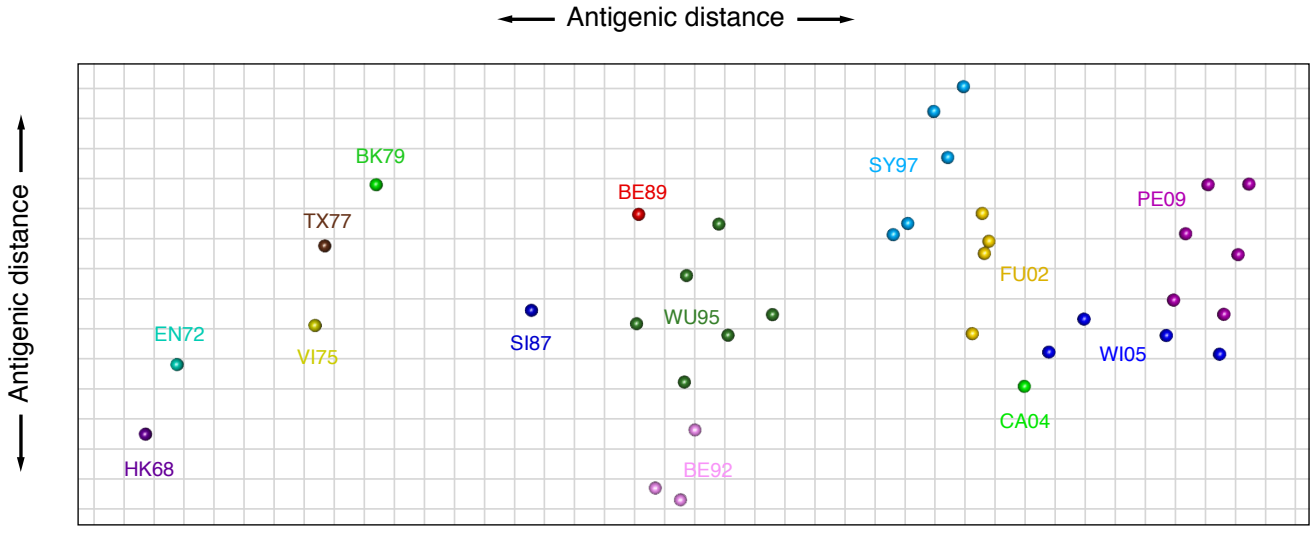



Figure S2: Antigenic map of the overlay-merge of the viruses used in this study, and the antigenic map as published in Smith *et al.* (*2*). Colored circles are viruses used for the human serum titrations, and color-coded by antigenic cluster (cluster name indicated). Viruses used to coordinate the map in the overlay-merging and not used further are shown as small black dots, sera were omitted for clarity. HK68: Hong Kong 1968; EN72: England 1972; VI75: Victoria 1975; TX77: Texas 1977; BK79: Bangkok 1979; SI87: Sichuan 1987; BE89: Beijing 1989; BE92: Beijing 1992; WU95: Wuhan 1995; SY97: Sydney 1997; FU02: Fujian 2002; CA04: California 2004; WI05: Wisconsin 2005; PE09: Perth 2009.

### 1.2.2 Titer correlation with antigenic distance

In order to investigate whether antigenic distance is, as for ferrets, related to HI titers in humans, we calculated the pairwise Spearman rank correlation of the HI titer responses between all virus pairs as a function of their pairwise antigenic distance (see Figure S3). Correlation coefficients decrease considerably with antigenic distance, from around 0.8 to below 0. This result clearly indicates that antigenically similar viruses have correspondingly similar HI titers, and thus that one can fit a surface through the HI data to create an antibody landscape. Correlations did

not reach one, even for the smallest antigenic distances, likely because of noise introduced through error in the HI assay and intrinsic differences between antigens not captured by this measurement of their antigenic characteristics. The negative correlation for large antigenic differences is likely a result of data structure, where young people have not been alive in certain decades, and are therefore more likely to be exposed to recent influenza viruses, and *vice versa* individuals getting less frequent infections in their adult years, yet having large titers against their childhood viruses.

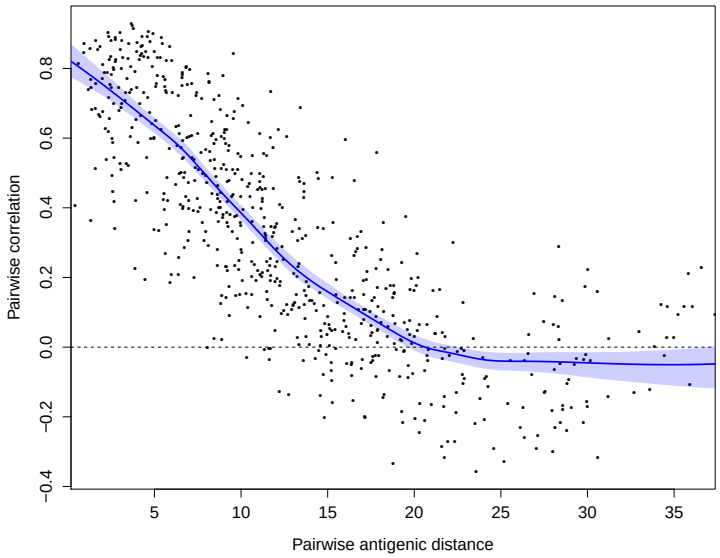



Figure S3: The antigenic distance between all possible antigen pairs is plotted against the corresponding Spearman's rank correlation coefficient of the corresponding set of HI titers for the two viruses in the pair, for all human samples. The solid blue line represents a Loess fit through the data (span 0.5), while the lighter blue regions indicate the 95% confidence interval for the same fit based on 500 bootstrap repeats, sampling with replacement.

### 1.2.3 Generating the antibody landscape

To generate the antibody landscape, titers were converted onto a log scale via:

$$z = log_2(\frac{\text{HI titer}}{10})$$

The log-converted HI titer is plotted as the z-value for each virus strain at the corresponding position on the antigenic map, a higher value corresponds to more antibody (Figure S4).

Each landscape was calculated using locally-weighted multiple linear regression based on the titration values available for the given serum sample against $n$ antigens. Each position on the antigenic map is therefore associated with three variables: the $x$ coordinate associated with the position in the antigenic map, $x_i$; the $y$ coordinate associated with the position in the antigenic map, $y_i$; and the converted HI titer, $z_i$. To determine the landscape value, a separate regression is performed for each point in the antigenic map.

For a given point $p_j$ (with associated antigenic coordinates $x_j$ and $y_j$), the predicted landscape height $\hat{z}_j$ (in log units) is calculated as follows. First, Euclidean distances are calculated to obtain the antigenic distance, $a_{ij}$, of each antigen with respect to the point in question, such that:

$$a_{ij} = \sqrt{((x_j - x_i)^2 + (y_j - y_i)^2)}$$

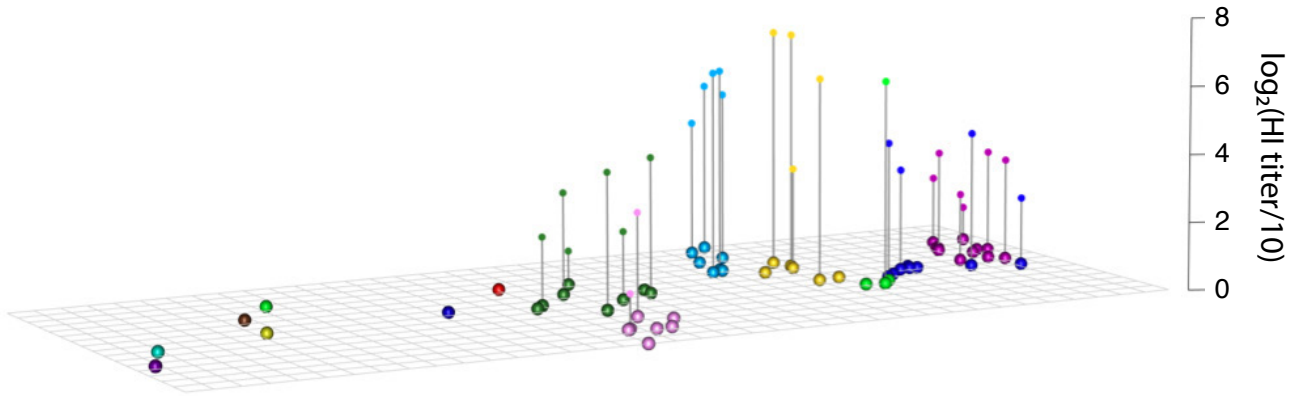



Figure S4: The antibody landscape is extended into an additional (3rd) dimension, to represent the log-transformed HI titer associated with each antigen when titrated against a given serum.

where $A$ is a variable that determines the maximum antigenic distance from $p_j$ to which titrations made against viruses will receive a non-zero weighting when calculating the landscape height (such that $2A$ effectively represents the bandwidth of the local regression). Multiple linear regression was then applied to characterize the relation between the measured titers $z_i$ and the coordinate variables $x$ and $y$, such that for each antigen:

$$z_i = c_j + x_i y_i \beta_j + \mathcal{E}_i$$

where the regression coefficients $c_j$ and $\beta_j$, which are different for each location in the antigenic map, due to the local nature of the regression, were resolved to minimize the weighted sum of squares of the errors (R function lm() used, with weights for each antigen point as defined.), $S$, given by:

$$S = \sum_{i=1}^{n} w_{ij} {\mathcal{E}_i}^2$$

Finally, predicted values for $p_j$ are calculated as follows:

$$\hat{z}_j = c_j + x_j y_j \beta_j$$

Because the HI assay starts at a dilution of 1:10, the minimum detectable antibody titer is 10. To properly model undetectable titers, we listed them as $\leq$5 in the HI assay or $\leq$-1 on the log scale and such values were allowed to take any value between -1 and -10 (because the value should be $\leq$-1, and is limited to aid optimization) when fitting the model. These estimated values for undetectable titers were calculated by minimizing the overall root-mean-square error (RMSE) of the model fit, which was optimized using a limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm (L-BFGS-B, iterating once from starting conditions of all non-detectable titers equal to -1). When calculating the RMSE, prediction errors were discounted when the predicted titer was below 0 and the measured titer was undetectable (in such a situation, we cannot assess the quality of the fit). When the predicted titer was above 0 and the measured titer was undetectable, the measured titer was assumed to have a value of -1 (as is frequently done for HI data, see e.g. Beyer *et al.* (*27*)). As an example, a predicted titer of 2 for a measured titer that was undetectable would be accorded an error of (2 - (-1)) = 3.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{j=1}^{n} (\hat{z}_j - z_j)^2}$$

where $z_j$ is the original titer value, and $\hat{z}_j$ is the predicted titer after fitting.

For the Ha Nam cohort data, the maximum detectable antibody titer was 640. Where the titer exceeded this value, and these are listed as $\geq$1280, the corresponding HI log titer is $\geq$7, and, equivalently to undetectable titers, values were allowed to take any value between 7 and 10 (the highest HI titer normally observed in an HI assay) and

optimized by minimizing the RMSE of the model fit.

By repeating this fitting procedure over a matrix of different antigenic coordinates, a landscape is plotted in three-dimensional space within the boundaries of the antigenic map. This surface shows higher antibody levels (higher HI titers) as elevations, and antigenic regions with lower antibody levels form depressions in the landscape (Figure S5).

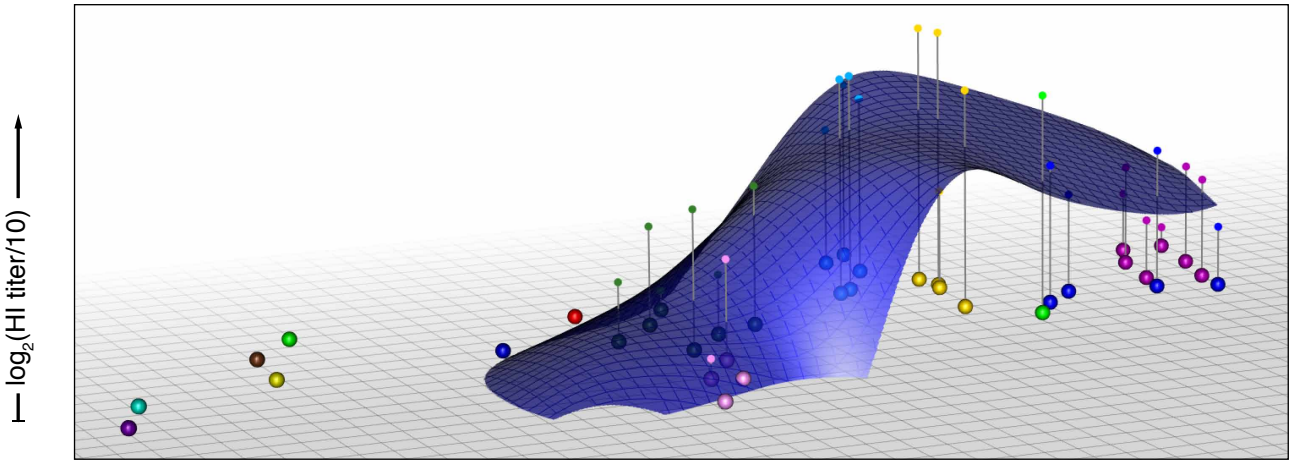



Figure S5: The landscape height is calculated for a matrix of antigenic coordinates within convex hull of the antigenic map, and the corresponding values show the antibody landscape.

### 1.2.4 Determining the potential for virus exposure

To approximate the likely extent of viruses to which an individual would have been exposed in their lifetime, we used information about the circulation of A/H3N2 antigenic clusters (see Figure 4 in Smith *et al.* (*2*) and Table S2), summarized in Table S9. To be conservative, individuals were considered exposed to any virus from the corresponding antigenic cluster if this cluster had circulated during any year from the year of birth to the year of sample collection.

| Antigenic cluster | HK68 | EN72 | VI75 | TX77 | BK79 | SI87 | BE89 | BE92 | WU95 | SY97 | FU02 | CA04 | WI05 | PE09 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Years circulating | 1968-1971 | 1972-1974 | 1975-1976 | 1975, 1977 | 1979-1986, 1988 | 1986-1991 | 1989-1992 | 1991-1996 | 1993-1997 | 1997-2004 | 2002-2005 | 2005-2006 | 2006-2009 | 2009-2012 |

Table S9: The years of circulation of the antigenic clusters.

### 1.2.5 Cross-validation

To determine the optimal value for parameter $A$ in the tricubic weighting function of the landscape fit, which governs the smoothness of the landscape by making each multiple linear regression more global (larger $A$ values) or local (smaller $A$ values), we used cross-validation. To represent all data sets (which varied in number and antigens against which serum was titrated), we performed 100 Monte-Carlo cross-validations on a random subset of 10 individuals from each of the 5 datasets. This subset of the data set was split into 500 training-test sets, with each training set consisting of a random subset of 75% of the available titrations (75% of the viruses measured for a given serum were retained, without replacement), and the test set comprising all sera and the remaining 25% of the viruses.

The training set was used to construct landscapes, and the predicted values of the landscape for the test set were compared with the actual HI results. The same 500 training-test set splits were then used to compare the prediction error for different settings of $A$. Subsequently, we repeated this procedure with 50% of the data in the training set, and 50% in the test set, using the same random subset of individuals.

To obtain the overall prediction error, the RMSE was calculated as described above and averaged across all 50 individual landscapes. If, in a given training-test set split, a test virus had fewer than three surrounding training virus titrations within six antigenic units (the smallest value of $A$ tested) that prediction value was excluded when calculating the RMSE for all values of $A$, and if the predicted and measured HI titer were both $<10$, this value was excluded when calculating the RMSE.

Based on the results shown in Figure S6 a value for $A$ of 11 was chosen, as this gave good prediction RMSE values, for both 75%-25% and 50%-50% training-test set splits, when compared to different values of $A$, and when compared to landscapes generated from randomized data (see section 1.2.7).

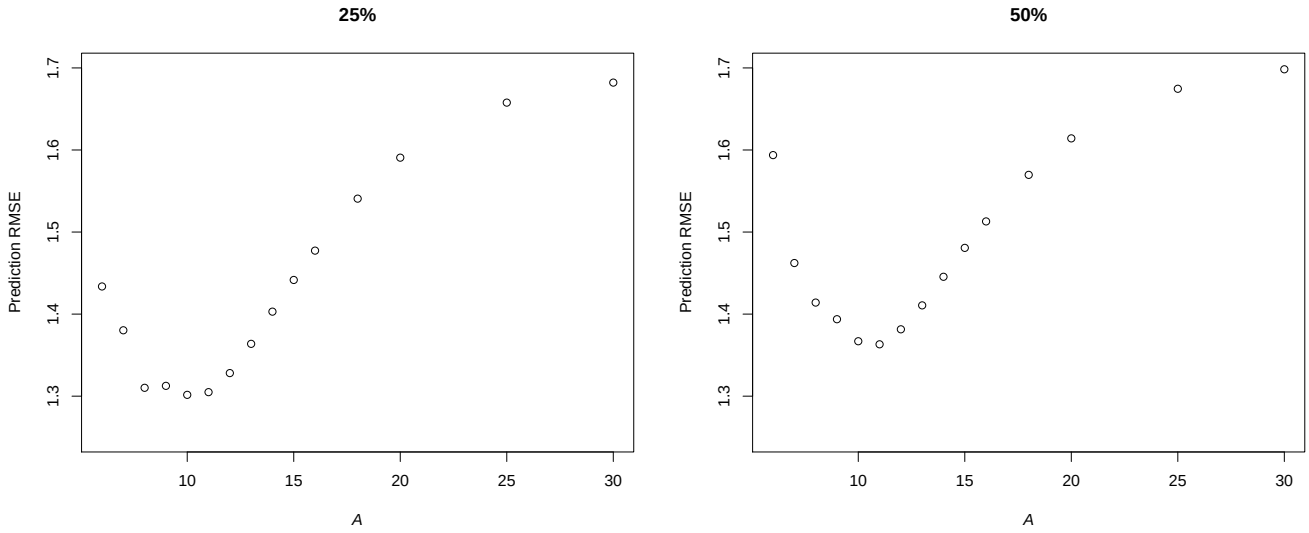



Figure S6: Root-mean-square error of prediction, with 100 cross-validation runs, for different settings of $A$ in the weighting function for locally-weighted multiple linear regression, leaving out 25% or 50% of the data.

In the 75%-25% split in Figure S6, a value of 8 also returned a low prediction RMSE, but is an outlier from the smooth trend of the other data points. We further investigated this data point and found that this value likely leads to over-fitting, which would lead to non-generalizable results: in Figure S7, for $A$ of 8, there are many features visible that cannot be supported by the data (e.g. the SI87 cluster in blue was only represented by 1 titration, but there are many narrow features in this region that seem spurious), see also 1.2.9.

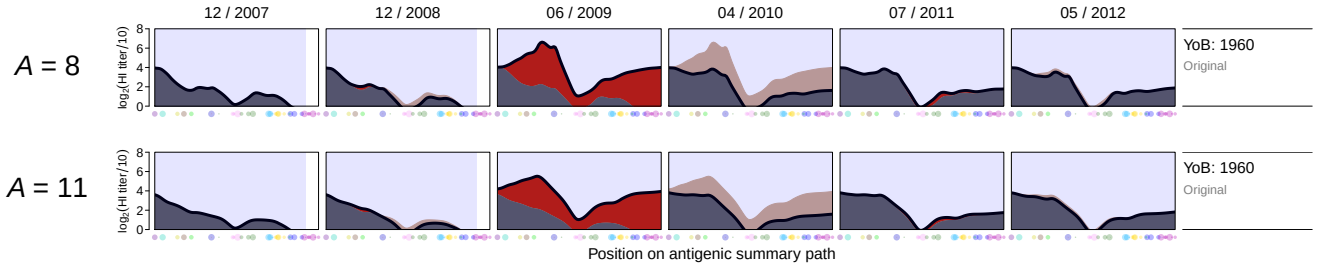



Figure S7: A comparison of landscape fits with when the smoothing parameters $A$ was 8 or 11, for an individual born in 1960 from the Ha Nam cohort.

### 1.2.6 Estimating HI accuracy

Inherent measurement error in the HI assay is well recognised, for example due to between-laboratory variation. In our study, all titrations were performed within the same laboratory, and so we were able to use the repeated samples taken from the same individuals in the Ha Nam cohort to quantify the repeatability of the results. Ideally, this would be done by measuring the error in repeated HI assay data, but such data was not available on a large scale. A more conservative estimate of repeatability may be obtained by investigating changes in antibody titers over consecutive years for individuals in the Ha Nam that were not infected (i.e. the individual did not seroconvert or have PCR-confirmed infection), assuming that these individuals have an approximately constant titer response.

We investigated the variability in HI titers between pairwise measurements for all samples of an individual who were not infected in any year (by PCR or seroconversion). Additionally, we removed individuals that had a response against any influenza strain of 16-fold or more, as this suggests serological changes, despite not fulfilling the seroconversion criterion.

Figure S8A shows the distribution of errors (difference in HI titer) between two repeat measurements of the same antigen, for the full set of pairwise combinations of HI "repeats" from the same individual in two subsequent years. Over 60% of the repeats show no change in the HI titer.

To further investigate whether the repeatability depends on the value of the titer, we calculated the error as a function of HI titer (based on the two entries from both titers forming the pair), see Figure S8B. The distribution of pairwise differences for HI titers of 0 and larger changed little, being distributed around a mode of 1, and was not significantly greater for greater HI values. When titers fell in the undetectable range ($<$10, corresponding to a log-value of -1 for this analysis), consistency between pairs of measurements was further improved, with a mode of 0 (where both measurements display undetectable cross-reactivity). This may be expected given the larger range of potential reactivity encompassed by an "undetectable" reading. To avoid artifically deflating the reported RMSE, such pairs of measurements where both titers were undetectable were excluded from the RMSE and MAE calculations in sections 1.2.5 and 1.2.7.

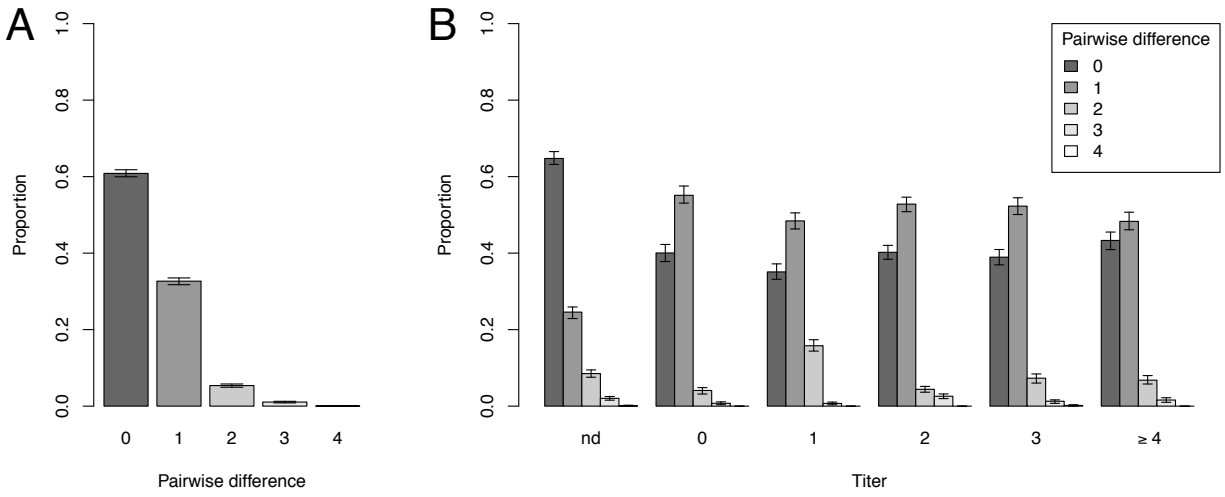



Figure S8: Estimates of HI consistency. (A) Proportional distribution of HI error across all "repeats". (B) Proportional distribution of HI error across "repeats" subset by the associated measured titer of either of the pair.

The RMSE for pairwise agreement between each "repeat"(with at least on of the titers is a numeric measurement, i.e. $\geq$10) is 0.90 (95% CI, 0.86 - 0.93). Note that there are very few instances of a pairwise difference of 4, providing evidence that the 4 individuals with a change of $>$4 had indeed most likely undergone a serological change unrelated to HI inaccuracy.

### 1.2.7 Estimating landscape accuracy

After parameterizing the model with $A = 11$, Monte-Carlo cross validation was used to determine the predictive ability of the model for the different data sets. Again, 100 repeats were performed using randomly generated 75%-25% training-test set splits. In addition, we randomly associated titrations with different viruses in the Ha Nam cohort study and used these shuffled values to construct landscapes, as a negative control. We subsequently calculated the root-mean-square error (RMSE) and the mean absolute error (MAE) for the predictions made for the test viruses in each individual landscape. Table S10 displays the mean of the RMSE values obtained across the different test-landscapes and its standard error, as well as the median RMSE and median absolute deviation (MAD); and the mean and standard deviation of the MAE values. If, in a given training-test set split, a test virus had fewer than three surrounding training virus titrations within six antigenic units, that prediction value was excluded from calculations. Also, if both the predicted and measured value were $<10$, this was not taken as an error of 0 but rather excluded from the analysis, because it would artificially deflate the RMSE value.

| Data set | mean RMSE (sd) | median RMSE (MAD) | mean MAE (sd) |
|---|---|---|---|
| Infection: Ha Nam cohort | 1.33 (0.42) | 1.29 (0.39) | 1.14 (0.38) |
| Pre-vaccination - 1997 study | 1.30 (0.45) | 1.26 (0.38) | 1.11 (0.40) |
| Post-vaccination - 1997 study | 1.33 (0.36) | 1.30 (0.33) | 1.10 (0.31) |
| Pre-vaccination - 1998 study | 1.28 (0.45) | 1.22 (0.38) | 1.10 (0.40) |
| Post-vaccination - 1998 study | 1.26 (0.36) | 1.22 (0.31) | 1.05 (0.31) |
| Randomized (Ha Nam) | 1.96 (0.69) | 1.87 (0.65) | 1.39 (0.62) |

Table S10: Prediction summaries for the different data sets. The root-mean-square-error (RMSE) and mean absolute error (MAE) were calculate for individual landscapes, and the distribution parameters (mean and standard deviation (sd), median and median absolute deviation (MAD)) for these results are displayed. To act as a negative control with which to compare RMSE values, we also calculated prediction errors when titrations were randomized across the antigens in the Ha Nam cohort (such that titrations are no longer related to their original antigen).

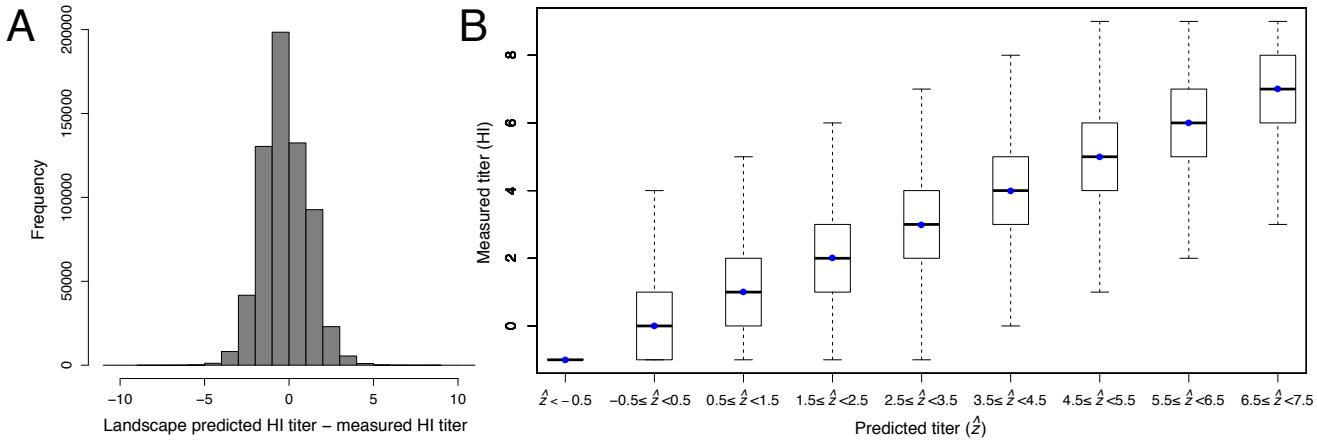



Figure S9: (A) Histogram of the prediction error (landscape predicted titer minus measured HI titer), across all data sets. (B) A series of boxplots showing the distribution of measured titers associated with predicted titers ($\hat{z}$) binned as indicated. The mid-point of each bin of predicted titers is plotted above it at x=y as a blue dot, whiskers extend to 1.5 times the inter-quartile range from the box.

The distribution of all errors across the cross-validation runs in the 4 vaccination data sets and the Ha Nam cohort data set is shown in Figure S9A; if both the measured and predicted HI titer were $<10$, this error of 0 was excluded. If, in a given training-test set split, a test virus had fewer than three surrounding training virus titrations within six

antigenic units, that prediction value was also excluded. The RMSE values in Table S10 for each cohort were similar and each substantially lower than the prediction error when viruses were only randomly associated with antigenic coordinates, indicating that the landscape does indeed model underlying structure in the dataset, and is not over-fitting the data. Given that the error of the HI assay is estimated to be around 0.9 (see section 1.2.6), the goodness of fit of the landscape is reasonable, but not perfect. Examining the distribution of errors associated with different fitted values (this time re-including titers that were measured to be $<10$) in Figure S9B, we can see that although there is error, measured titers distribute evenly around the predicted value at the full range of landscape values found in this study (the highest landscape peak across all our studies was 7.42).

Figure S10 displays the distribution of RMSE per individual (i.e. across the 100 training-test set repeat predictions). It shows that individuals are approximately equally well represented by a landscape, and the errors are of a similar size as when calculated across the data set as a whole (see Table S10).

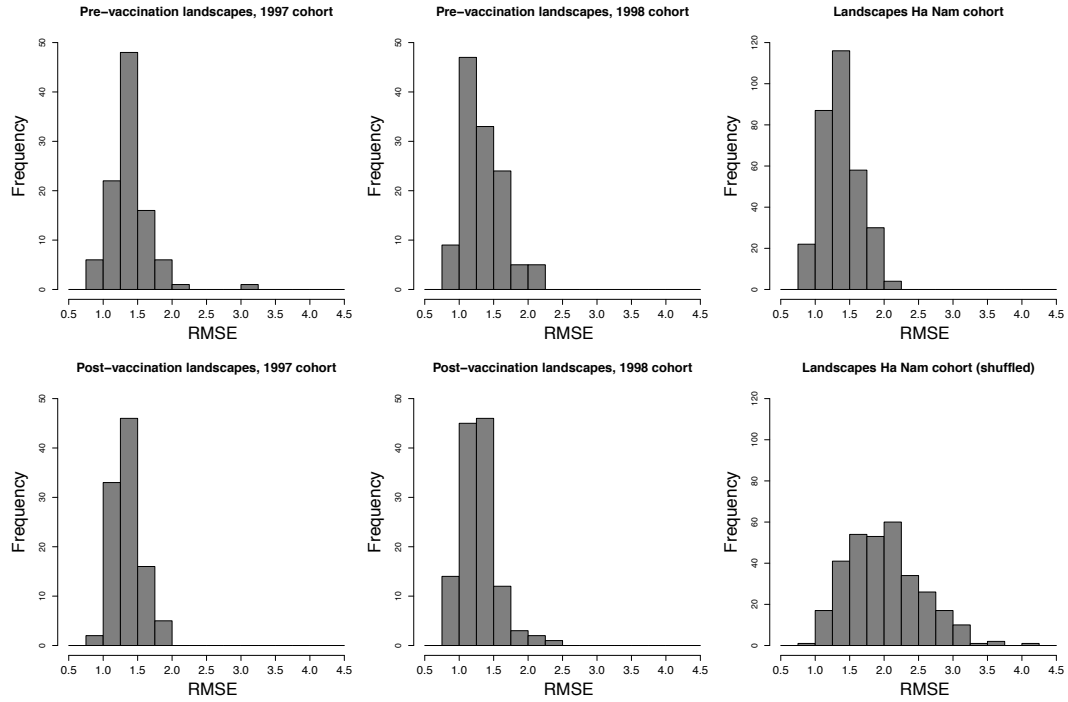



Figure S10: Histograms of the root-mean-squared-error (RMSE) of prediction calculated per individual, for the 4 vaccination data sets, the Ha Nam cohort, and the negative control (Ha Nam cohort shuffled).

Finally, we also wished to investigate whether there were any virus-specific differences, e.g. if some viruses are consistently under- or over-predicted in the landscape. Figure S11 shows histograms of prediction errors per virus for the Ha Nam cohort and the vaccine data sets. These figures indeed demonstrate a consistent bias in prediction accuracy of viruses, for example NL/620/89 is consistently under-predicted in both data sets, and Victoria/361/11 is consistently predicted to be higher than it was measured. A similar phenomenon has been described previously by Lessler *et al.* (*5*).

This phenomenon of virus-specific differences in reactivity is the source of the relatively larger RMSE errors we report above, and related to the biological limit of the correlation in section 1.2.2 to 0.8, instead of 1. These difference may be in part due to inaccuracies in the representation of antigenic differences between viral strains, and to factors unrelated to antigenic characteristics, such as intrinsic differences between the avidity of antibody binding. Such intrinsic differences, which may bias conclusions when investigating only a single virus at a time, are overcome by the antibody landscapes methodology, as a selection of viral strains are used to infer overall antibody reactivity.

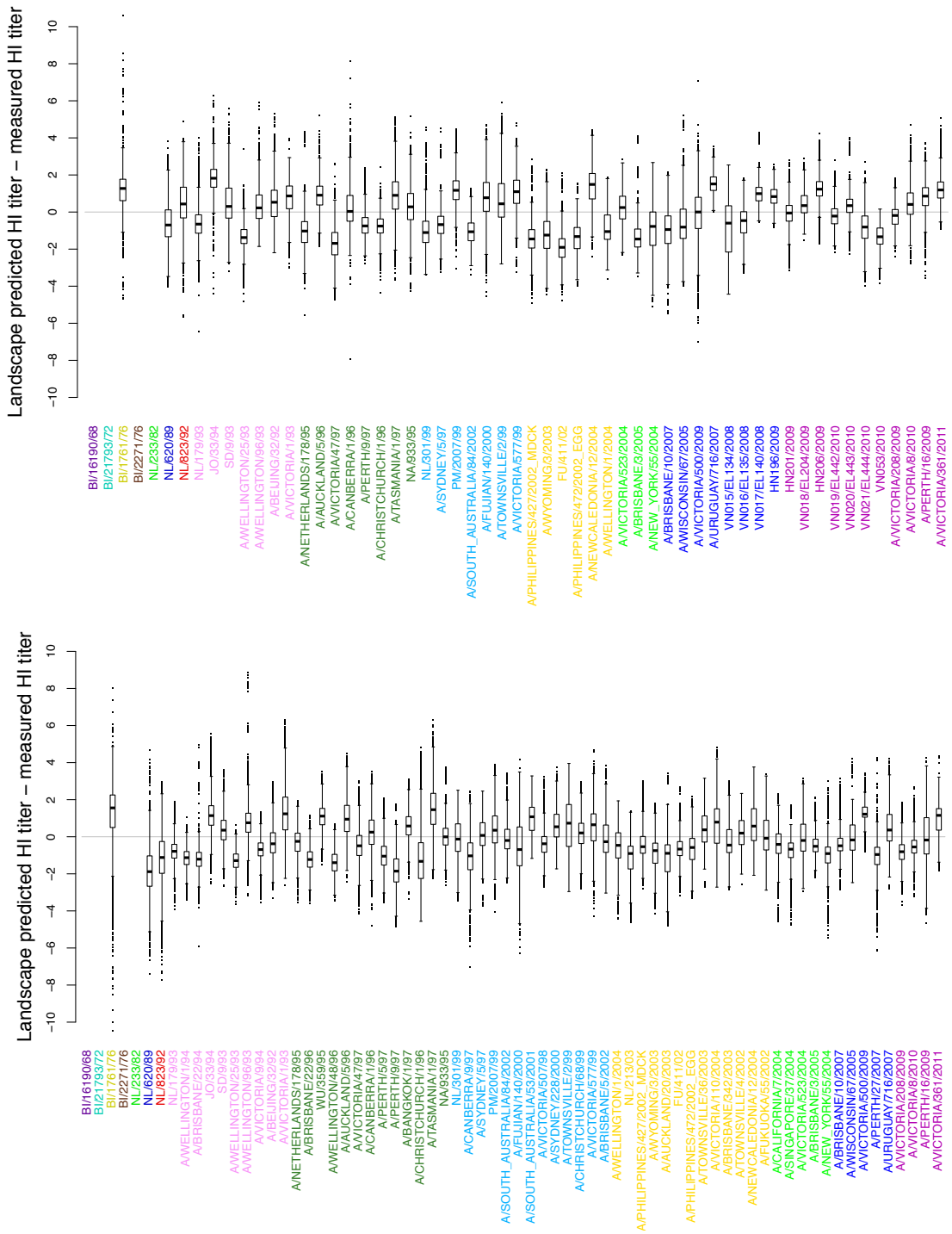



Figure S11: Histograms showing distribution of prediction errors (calculated as predicted HI titer minus measured HI titer) per virus, calculated across all individuals and repeats. Top: based on the Ha Nam cohort data. Bottom: based on the vaccine data sets (pre- and post-vaccination landscapes for the 1997 and 1998 cohorts). Titers predicted to be $<10$ were excluded from the data set when the measured HI titer was -1, and if a test virus had fewer than three surrounding training virus titrations within six antigenic units, that prediction value was excluded from calculation, hence some viruses in sparse regions of the antigenic map have no values associated with them.

### 1.2.8 Calculating a summary path

To create a rotation-independent visualization of the three-dimensional landscape, and additionally to facilitate comparison of multiple antibody landscapes, we chose to summarize the landscapes by taking a path that passes through the different antigenic clusters, thus covering the relevant regions of antigenic space. Figure S12 shows the result generated by fitting a smoothing spline (smoothing parameter 0.1, Matlab R2013, Natick, MA) through the antigenic positions of the viruses used in this study, and all viruses except those of the cluster BE89 (an evolutionary dead-end) that were included in the map by Smith *et al.* (*2*).

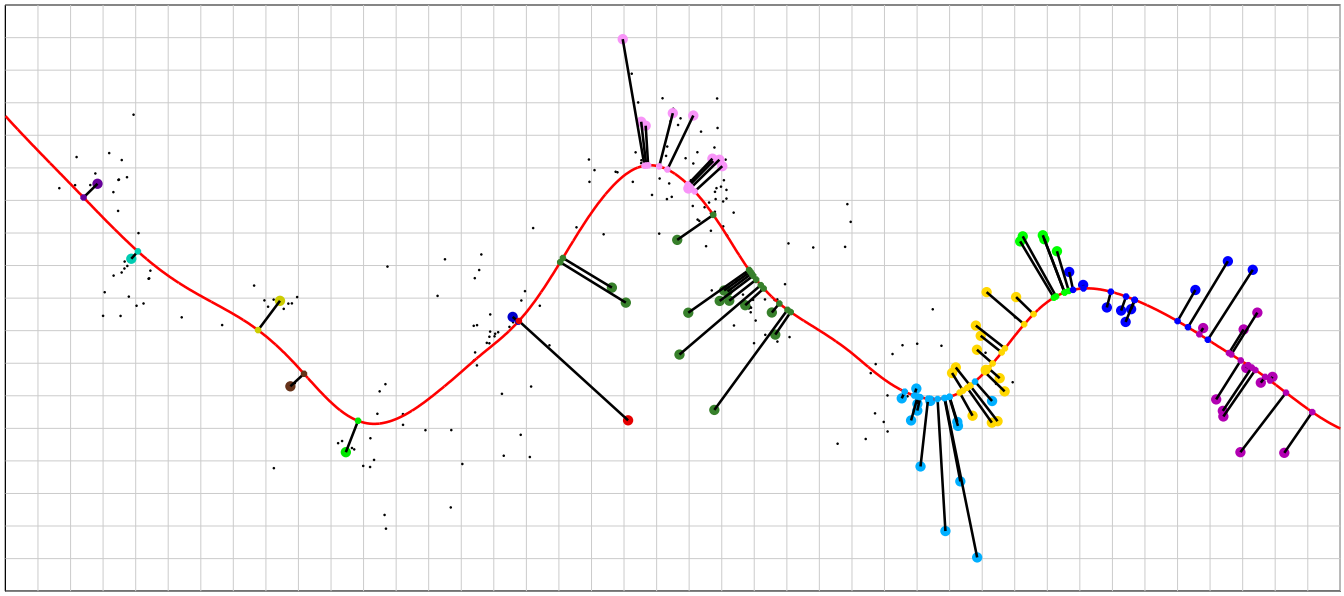


Figure S12: The smoothing spline through the antigenic map data, and the corresponding projections of the viruses used in our studies onto this path.

When evaluating the summary representations along the path, it should be kept in mind that the axis of this path does not equal antigenic distance, because the path curves, and therefore two points along the path may be closer or further than their direct, Euclidean distance calculated from the respective two-dimensional antigenic coordinates.

### 1.2.9 Bootstrapping the viruses used to generate the landscapes

Antibody landscapes are necessarily an abstraction of the raw HI titers, to enable both visual analysis and comparison of large amounts of data. The combination of multiple HI measurements may reduce the error associated with a single HI measurement, and thus increase the accuracy with which antibody levels can be inferred. However, there is variation in HI titers between viruses that is not captured by the measurements of antigenic similarity we use and as such, some details are lost when viewing results in an antibody landscape. By using many viruses however, the presentation of serum antibody reactivity achieved by antibody landscapes is one that is more robust than serological results considered in isolation.

Figure S13 is an example to demonstrate a visualization of the degree of certainty with which one can interpret the features of the landscape, here for the individuals from Figure 2 in the manuscript. By sampling with replacement from the set of viruses, we created 500 bootstrap repeats, and calculated the antibody landscapes for each bootstrapped data set. The resulting confidence intervals give some indication of the variability that could be expected from changes in e.g. the selection of antigenically different viruses, and their influence on the landscape shape. Based on the variability seen in this figure, and also the consistency of the landscape shape of different individuals over time, we are confident to interpret the global landscape shape when inferring patterns of the "true landscape", but more hesitant to look at narrow features - further supporting the choice of value 11 for parameter $A$ in section

1.2.5. In particular, the regions of the landscape that are more sparsely sampled (e.g. in the older antigenic clusters) have less information, and must therefore be more cautiously interpreted.

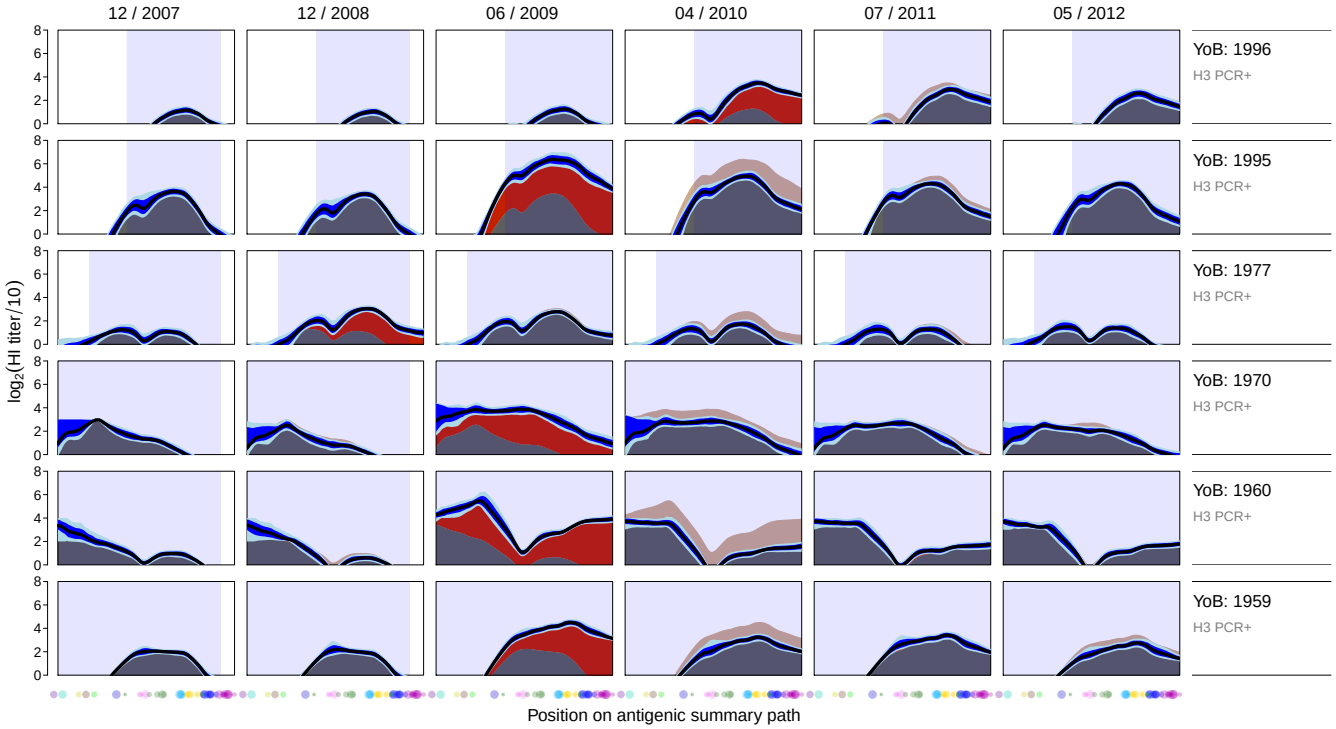



Figure S13: The landscapes from Figure 2 of the main manuscript of PCR-confirmed A/H3N2 infected individuals from the Ha Nam cohort (gray area, solid line) were subjected to bootstrap resampling (500 draws of the viruses, with replacement). As a result of the bootstrapping, confidence intervals of the landscape shape were obtained and displayed with respect to the landscape shape (grey, increase and decrease with respect to the previous year shown in red and beige, respectively), and showing the 50% (dark blue area) and 75% (light blue area) confidence intervals of the landscape shape.

### 1.2.10 Sensitivity of antibody landscape to number of viruses

Titrations on the scale undertaken in this study are rarely performed, and time-consuming. For the purposes of future studies, we therefore investigated if and how a more limited set of titrations would affect the calculated antibody landscapes. Figure S14 shows the landscapes calculated for individuals in the two vaccination studies using either the initial data set of 70 viruses, and a subset of 20 viruses chosen such that they were the approximately evenly distributed throughout the original antigenic space (determined with a Kennard-Stone algorithm (*28*)).

From this comparison, it can be seen that, although the details of the landscape change, these changes are within the range of uncertainty of the landscape as was illustrated in Figure S13, and the conclusions about the effects and differences between the two vaccines remain unaffected.

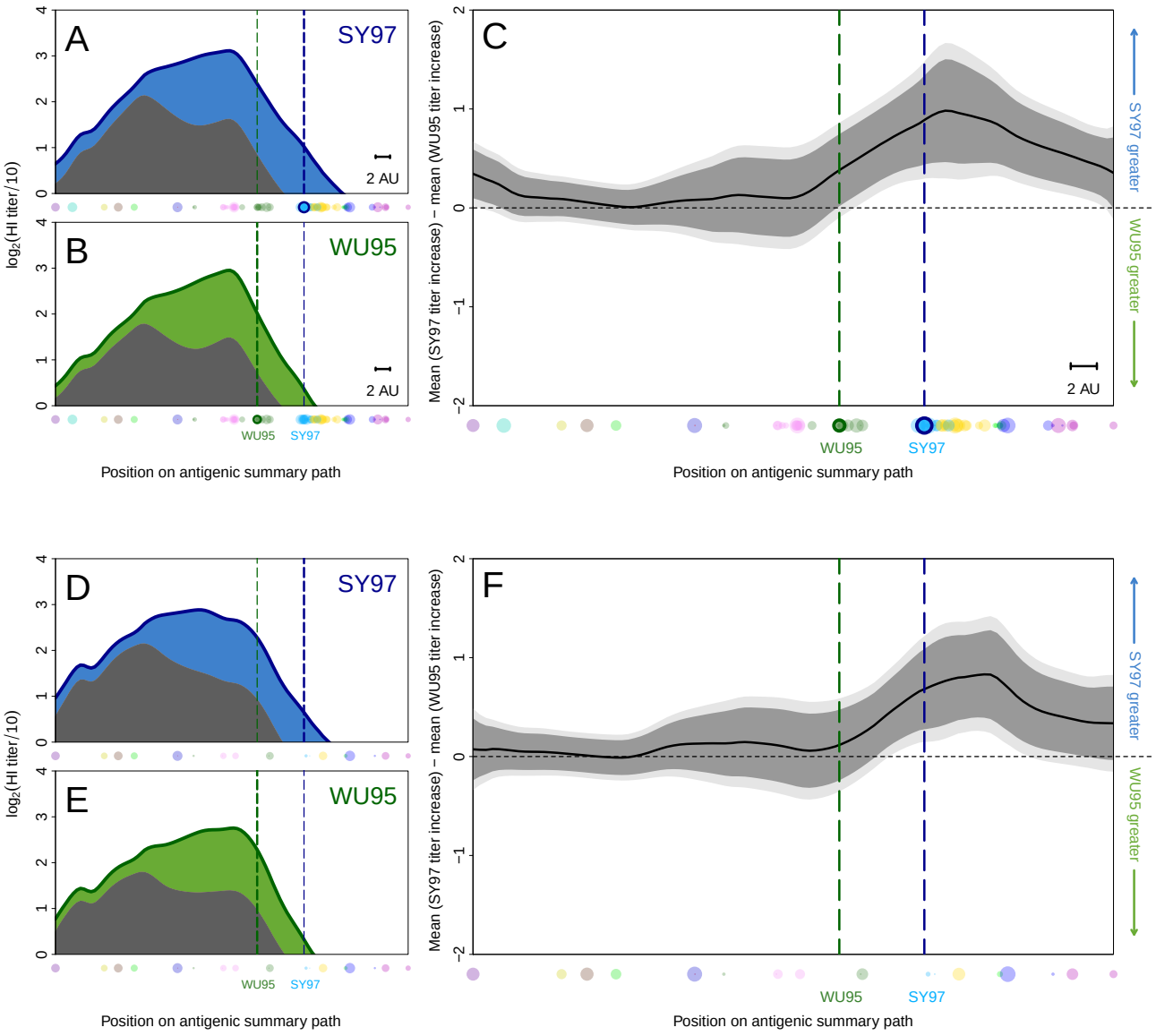
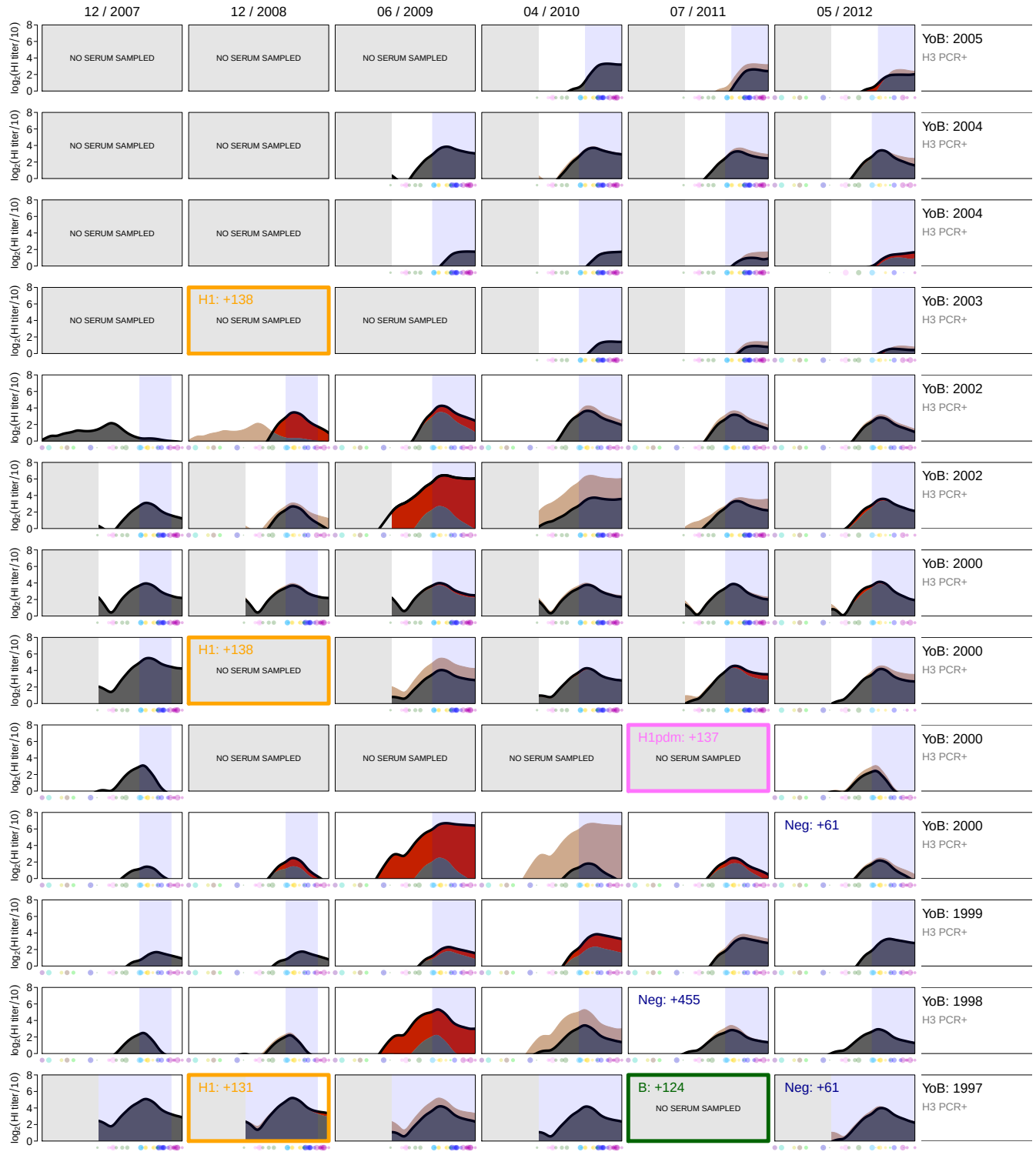


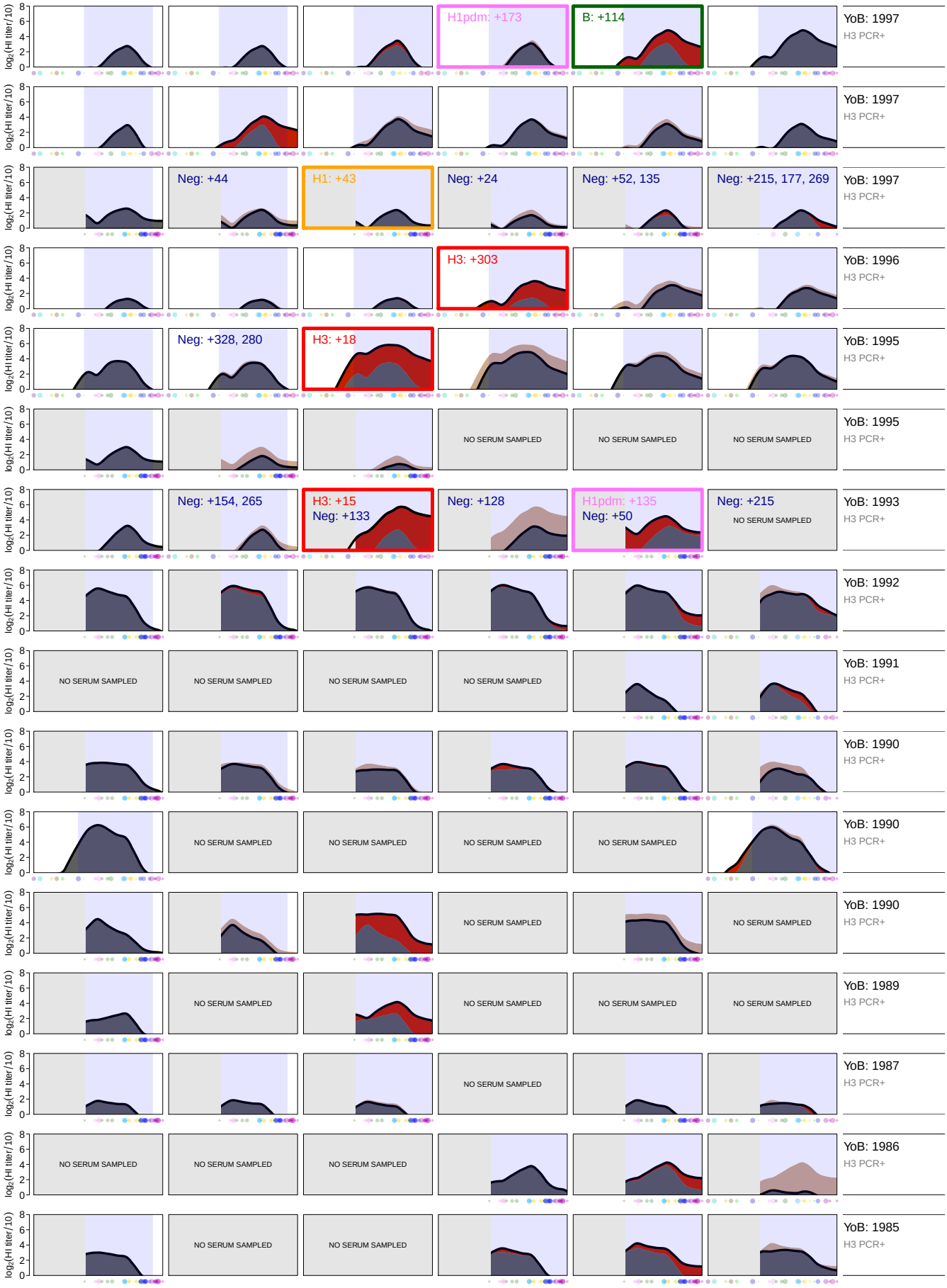

Figure S14: (A-C) The antibody landscape results for the vaccination studies based on titration against 70 viruses, as shown in the manuscript Figure 3. (D-F) The antibody landscape results for the vaccination studies when based on a selection of 20 viruses.

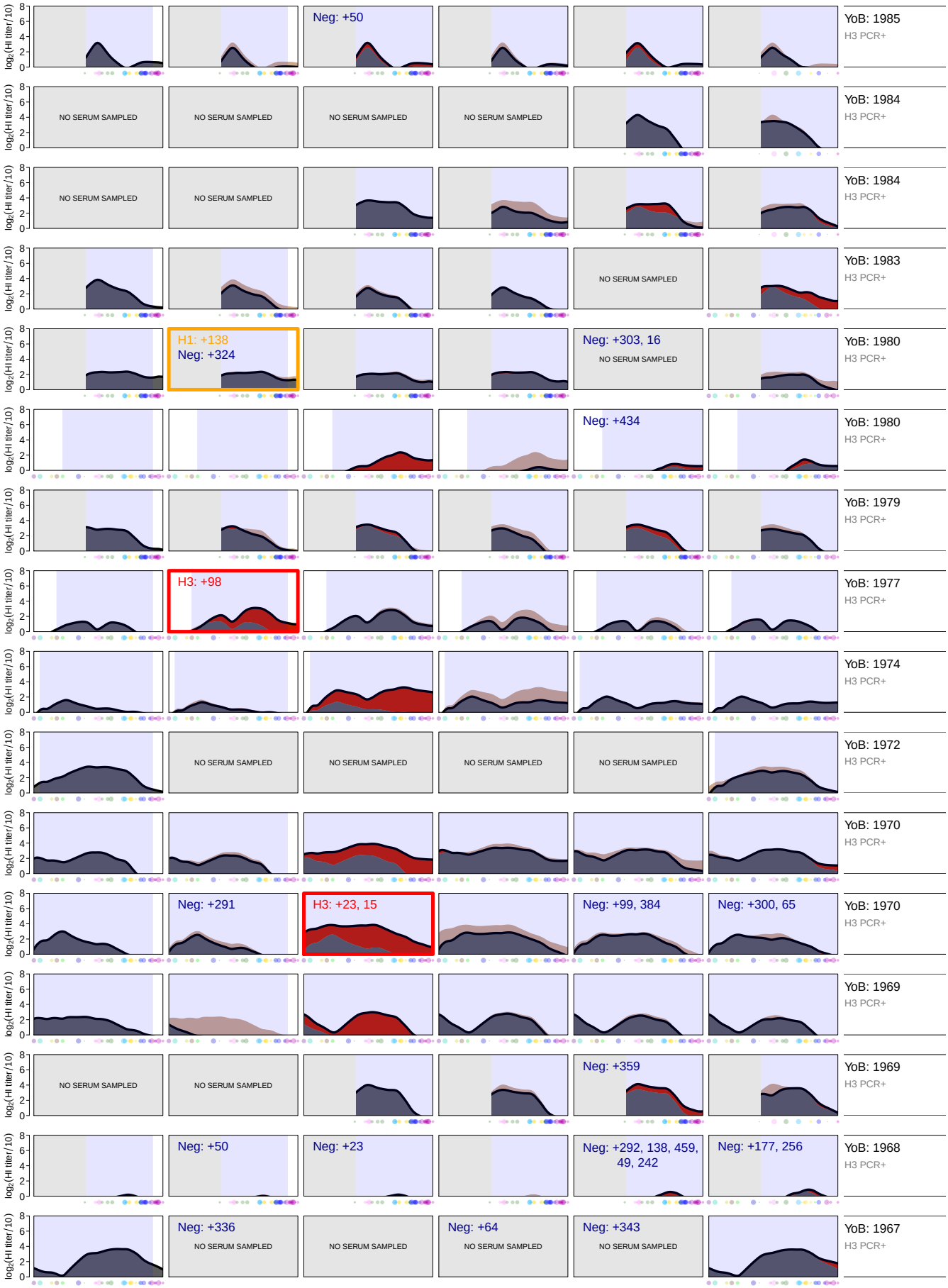
# 2 Response to infection

## 2.1 Plots of individual landscapes

Figure S15 shows the antibody landscapes for 2007-2012, where available, for all 69 individuals in the Ha Nam cohort study, sorted by year of birth (YoB).




log2(HI titer/10)
H1pdm: +173
B: +114
YoB: 1997
H3 PCR+
YoB: 1997
H3 PCR+
Neg: +44
H1: +43
Neg: +24
Neg: +52, 135
Neg: +215, 177, 269
YoB: 1997
H3 PCR+
H3: +303
YoB: 1996
H3 PCR+
Neg: +328, 280
H3: +18
YoB: 1995
H3 PCR+
NO SERUM SAMPLED
YoB: 1995
H3 PCR+
Neg: +154, 265
H3: +15
Neg: +133
Neg: +128
H1pdm: +135
Neg: +50
Neg: +215
NO SERUM SAMPLED
YoB: 1993
H3 PCR+
YoB: 1992
H3 PCR+
YoB: 1991
H3 PCR+
YoB: 1990
H3 PCR+
YoB: 1990
H3 PCR+
YoB: 1990
H3 PCR+
YoB: 1989
H3 PCR+
YoB: 1987
H3 PCR+
YoB: 1986
H3 PCR+
YoB: 1985
H3 PCR+


log2(HI titer/10)
Neg: +50
YoB: 1985
H3 PCR+
NO SERUM SAMPLED
YoB: 1984
H3 PCR+
NO SERUM SAMPLED
YoB: 1984
H3 PCR+
NO SERUM SAMPLED
YoB: 1983
H3 PCR+
H1: +138
Neg: +324
Neg: +303, 16
NO SERUM SAMPLED
YoB: 1980
H3 PCR+
Neg: +434
YoB: 1980
H3 PCR+
YoB: 1979
H3 PCR+
H3: +98
YoB: 1977
H3 PCR+
YoB: 1974
H3 PCR+
NO SERUM SAMPLED
YoB: 1972
H3 PCR+
YoB: 1970
H3 PCR+
Neg: +291
H3: +23, 15
Neg: +99, 384
Neg: +300, 65
YoB: 1970
H3 PCR+
YoB: 1969
H3 PCR+
NO SERUM SAMPLED
Neg: +359
YoB: 1969
H3 PCR+
Neg: +50
Neg: +23
Neg: +292, 138, 459, 49, 242
Neg: +177, 256
YoB: 1968
H3 PCR+
Neg: +336
NO SERUM SAMPLED
Neg: +64
NO SERUM SAMPLED
Neg: +343
NO SERUM SAMPLED
YoB: 1967
H3 PCR+

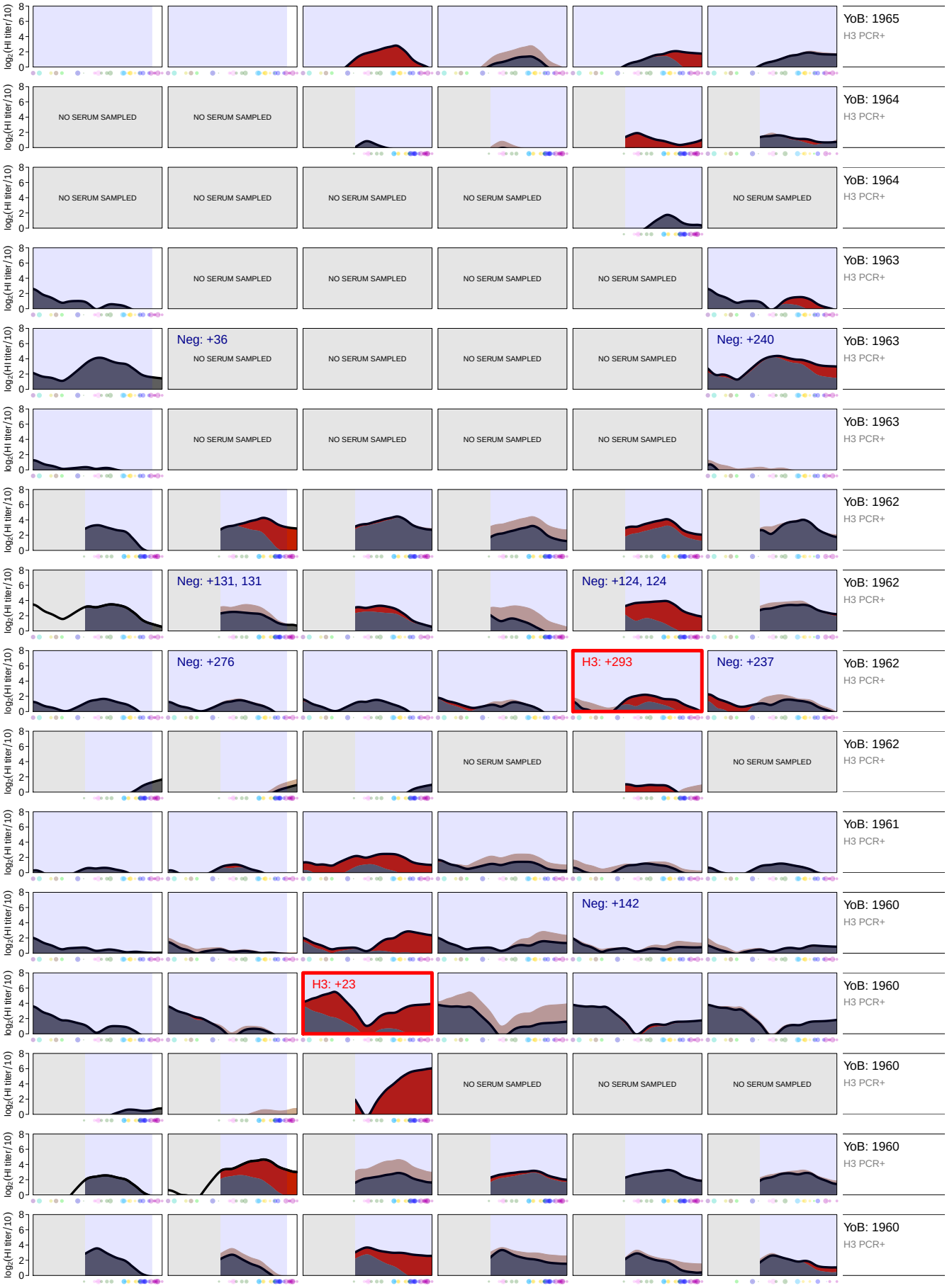

log2(HI titer/10)
YoB: 1965
H3 PCR+
NO SERUM SAMPLED
YoB: 1964
H3 PCR+
YoB: 1964
H3 PCR+
YoB: 1963
H3 PCR+
Neg: +36
Neg: +240
YoB: 1963
H3 PCR+
YoB: 1963
H3 PCR+
YoB: 1962
H3 PCR+
Neg: +131, 131
Neg: +124, 124
YoB: 1962
H3 PCR+
Neg: +276
H3: +293
Neg: +237
YoB: 1962
H3 PCR+
YoB: 1962
H3 PCR+
YoB: 1961
H3 PCR+
Neg: +142
YoB: 1960
H3 PCR+
H3: +23
YoB: 1960
H3 PCR+
YoB: 1960
H3 PCR+
YoB: 1960
H3 PCR+


Figure S15: (see above, pages 20 to 24) Antibody landscapes for all individuals in the Ha Nam cohort, from the original selection of households and individuals selected as PCR-confirmed infection, sero-conversion or control. The black line represents the landscape height for a given year along the path (between 2007-2012, per row). The red shading indicates an increase compared to the previous sample and beige a decrease. The blue-shaded area indicates antigenic clusters that circulated during an individual's lifespan until sample collection. The first sample after a PCR-confirmed influenza virus infection is marked with a colored box, and the corresponding colored number gives the days after ILI: red corresponds to a sample PCR-positive for A/H3N2, orange for A/H1N1, pink for A/H1N1(pdm), green for influenza B. Samples that were PCR negative for any influenza subtype are indicated with a blue box. Titrated virus strains are shown in their corresponding positions along the x-axis, symbol radius is inversely proportional to antigenic distance from the path, symbol color indicates antigenic cluster. Contemporary viruses likely having caused the infection are indicated with a red horizontal bar. The landscapes were cropped based on the most limited set of titrations performed for a given individual, the cropped area is indicated in gray (the gray line indicates the extended landscape in years where there were available titrations outside the cropped area). Landscapes marked with * were identified as possible sample swaps (see section 2.2.2).

## 2.2 Analyses of infection data

### 2.2.1 Quantitative measurement of landscape correlation

We noted a large degree of variability between landscapes of different individuals, yet for a given individual, the landscape shape was largely maintained over time. To quantify the individuality of the antibody landscape shape of a person, we compared the Pearson correlation of all pairwise antibody landscapes (summary along the path) from samples from the same individual in different years versus the correlation with landscapes constructed from samples taken from other individuals in all years. Average within-person correlation was $0.86 \pm 0.22$ compared to an average between-person correlation of $0.28 \pm 0.21$ (t-test, p-value $\ll 0.0001$). Figure S16A shows the between-individual correlations in black boxplots, and red dots are superimposed showing the within-individual correlations. Figure S16B shows this analysis without individuals with PCR-confirmed A/H3N2 infection or seroconversion in any year, for which it is expected that the landscape would change. For this subset, the mean within-person correlation was $0.89 \pm 0.25$, the mean between-person correlation was $0.26 \pm 0.23$ (t-test, p-value $\ll 0.0001$).

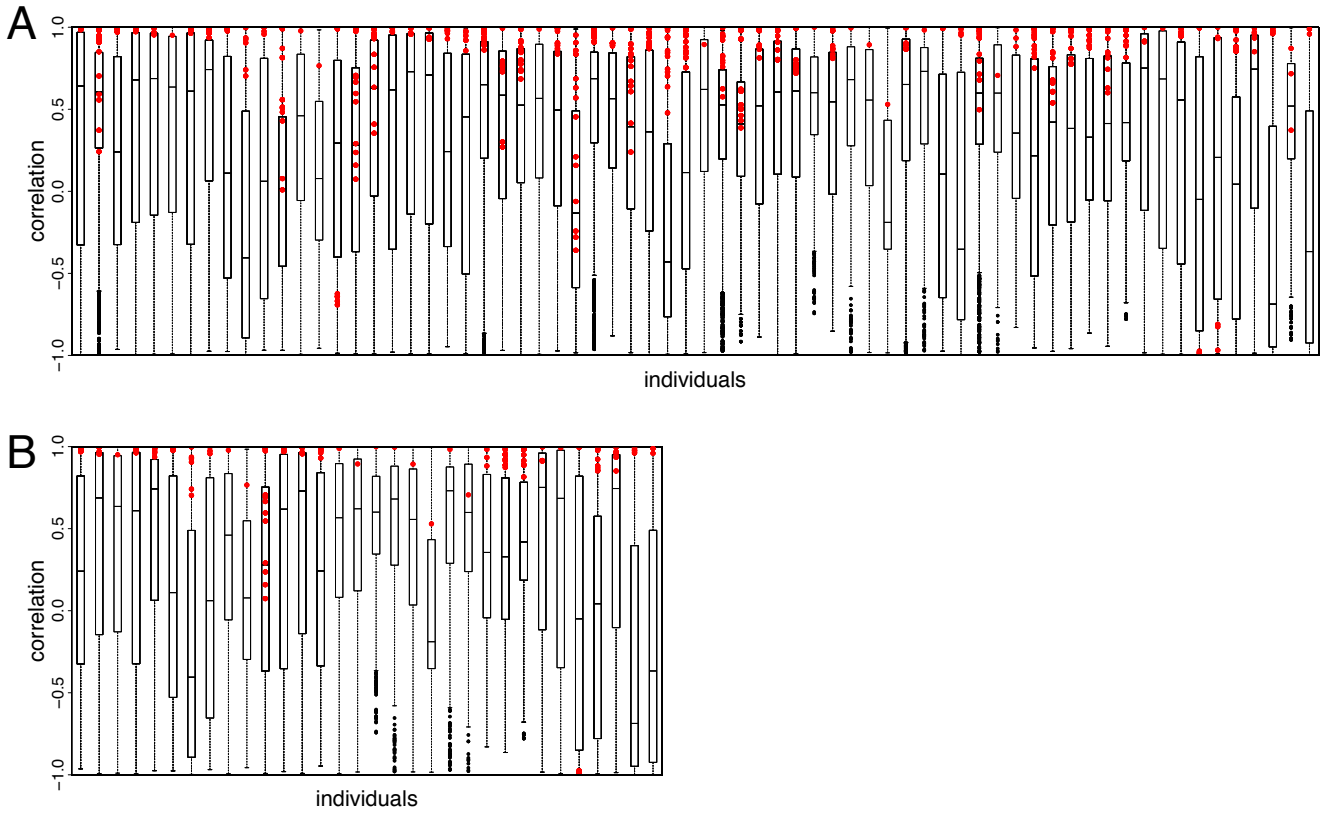



Figure S16: Boxplots showing the between-person correlation of antibody landscapes in black, and the within-person correlation of antibody landscapes as red dots (25% and 75% percentiles are the box limits, the whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box). (A) For all individuals for which landscapes were available for more than one year. (B) For the subset of individuals that had no PCR-confirmed infection or seroconversion in any year.

### 2.2.2 Evaluating potential sample swaps

Serum samples were collected, labeled, and transported carefully but inevitably errors occur, which can be difficult, if not impossible, to trace. Guided by the unusually low within-person correlations for a small number of individuals, in contrast to the usually high within-person correlations, we identified likely sample swaps based on the appearance of landscapes (Figure S17, Figure S18, Figure S19 and Figure S20).

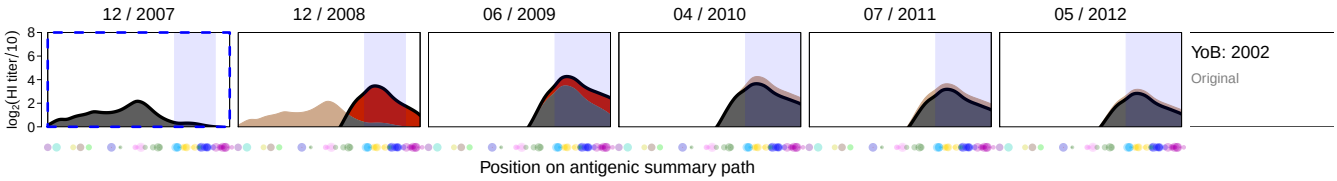
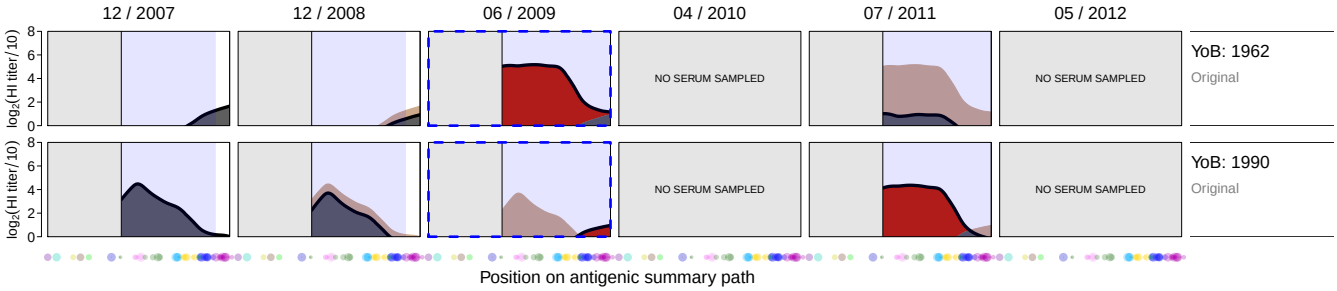



Figure S17: A potential sample swap in 2007 (legend as Figure S15).

Original:




Switched:

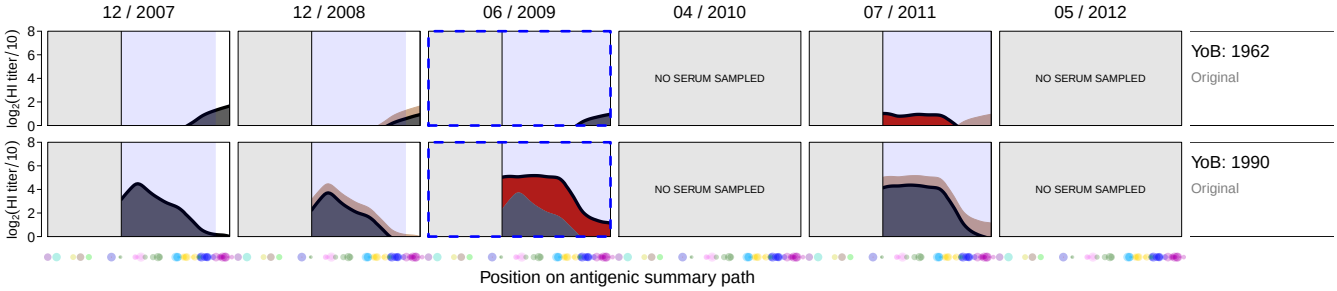
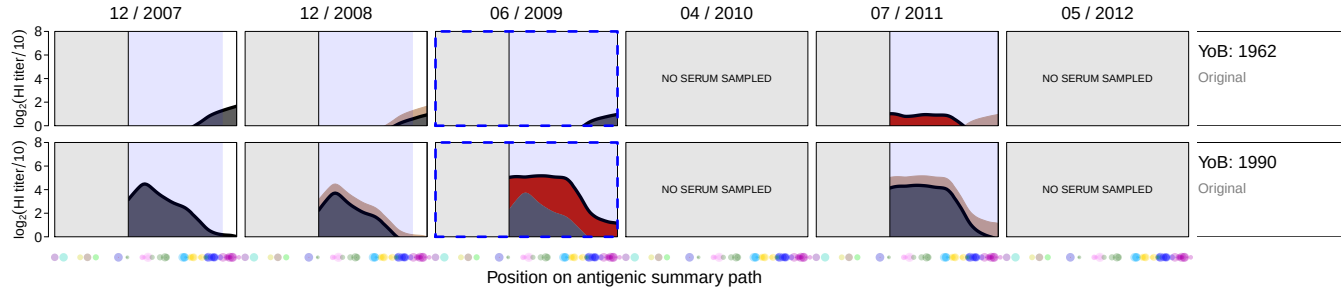



Figure S18: A potential sample swap in 2009 for individuals within the same household (legend as Figure S15).

To aid interpretation, where landscape series are shown we have switched back all the likely sample swaps shown above except the individual born in 2002 (Figure S17, since the likely pair of the swap could not be determined), and these samples are indicated with an asterisk. We also did this switching back of likely sample swaps when defining and showing seroconverters.

Original:




Switched:

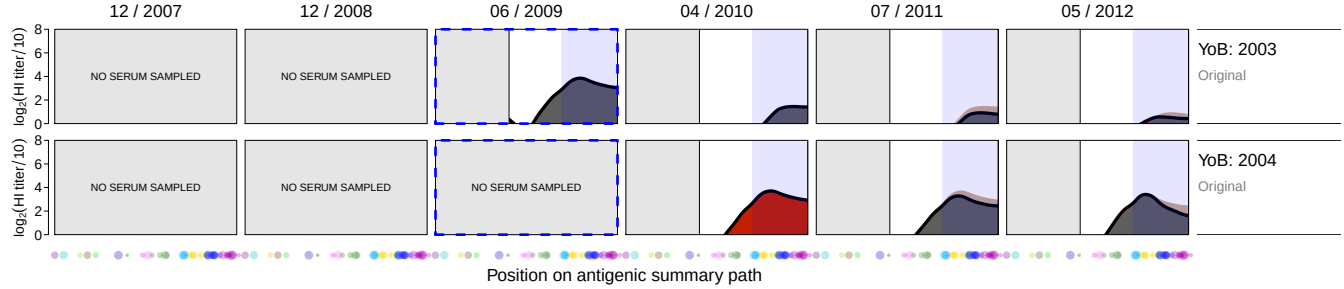
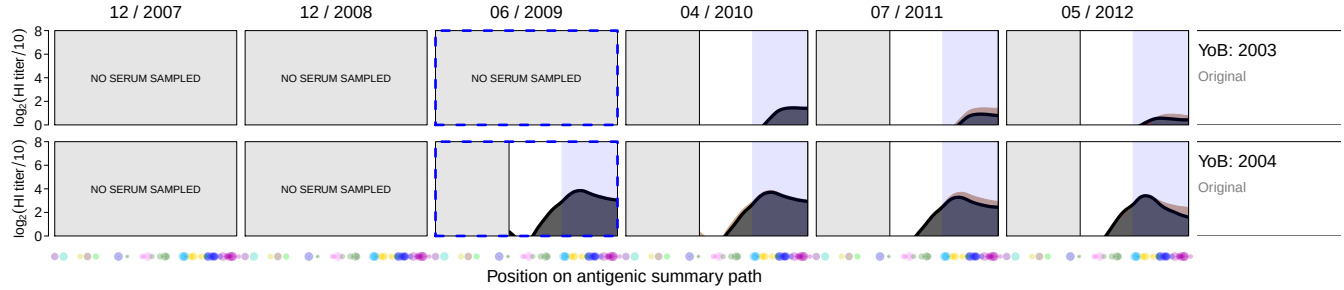



Figure S19: A potential sample swap in 2009 for individuals within the same household (legend as Figure S15).

Original:




Switched:

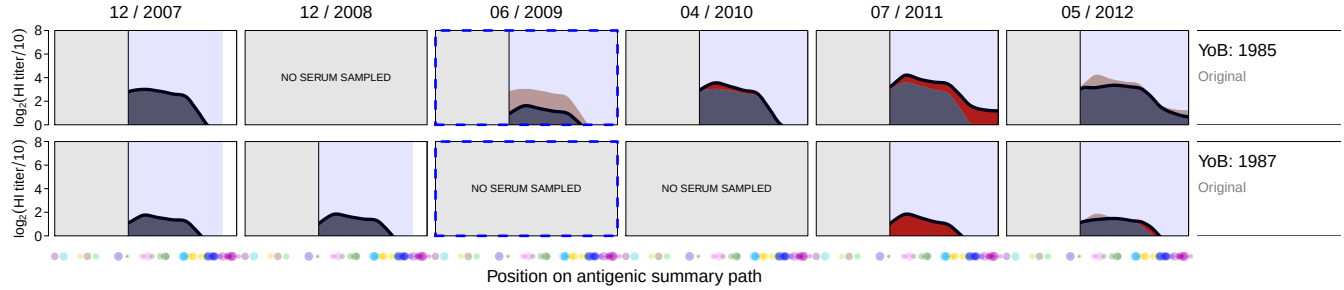



Figure S20: A potential sample swap in 2009 for individuals within the same household (legend as Figure S15).

### 2.2.3 The extent of cross-reactivity to viruses circulating before birth

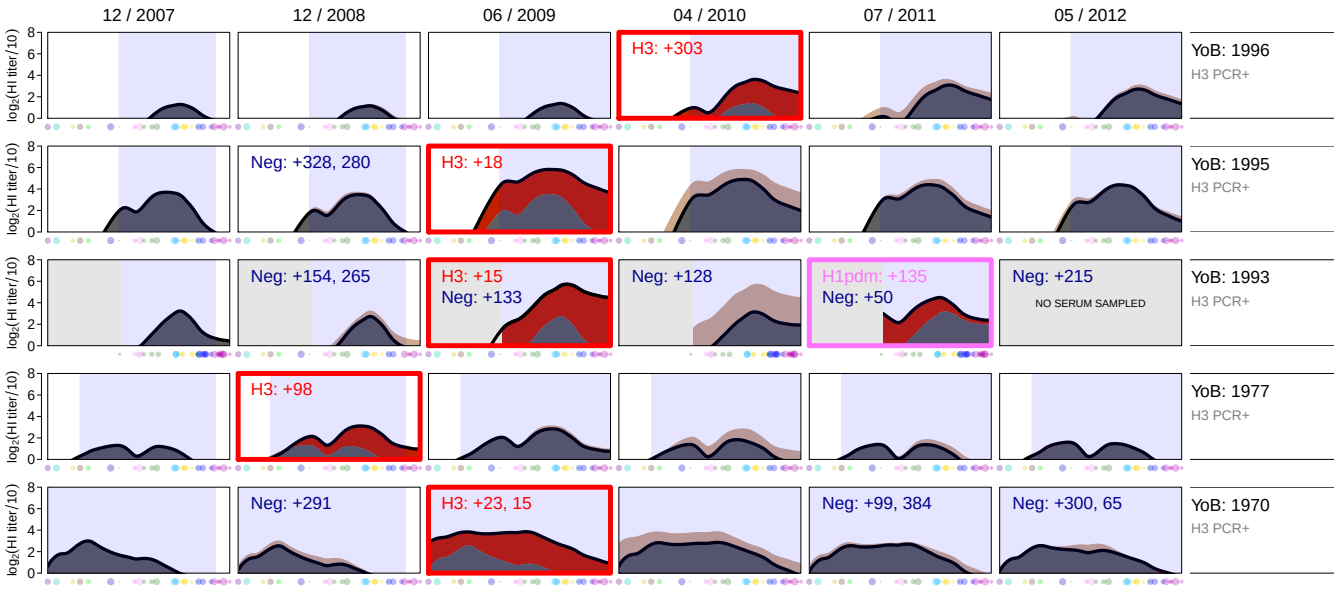
To provide an estimate of the typical extent of cross-reactivity of human antibodies, we investigated HI titers present against viruses from clusters circulating prior to an individual's birth, as these were most likely a result of cross-reactivity from antibodies formed against later clusters. However, we needed to control for singular titrations that recorded some reactivity for a virus (where the true signal is difficult to differentiate from the measurement error), and avoid changes due to a likely sample swap. Thus, we determined the lowest measured titer over all the samples from an individual per virus, and calculated how many antigenic clusters prior to the cluster circulating at birth included a virus that consistently gave measurable titers. Because cross-reactivity cannot be measured if the first measurable titer is of the first antigenic cluster that the individual was titrated against, we only included individuals where titrations extended beyond the antigenic range of measured titers, ensuring that the limit of cross-reactivity fell within the antigenic range of titrations. Additionally, individuals with no detectable titers to the cluster circulating in the year they were born were excluded, as they were possibly not infected by influenza from that cluster. The cross-reactivity span for these individuals (n=17) is shown in Table S11. Thus, the cross-reactivity of antibodies appears limited to a maximum of 2 antigenic clusters.

| Number of clusters prior to birth with detectable cross-reactivity | Number of individuals |
|---|---|
| 0 | 7 |
| 1 | 7 |
| 2 | 3 |
| 3 | 0 |

Table S11: The number of antigenic clusters relative to the antigenic cluster circulating at birth, for which antibody levels were consistently measured for an individual, in order to estimate the extent of antibody cross-reactivity of antibodies that is maintained long-term.

### 2.2.4 The antibody response in individuals with PCR-confirmed A/H3N2 infection

Figure S21 shows the antibody landscapes for 2007-2012, where available, for the 12 individuals in our study who reported ILI that was subsequently PCR-confirmed to be an A/H3N2 infection. We noted an antigenically broad response that commonly extended to even the oldest A/H3N2 viruses, although the breadth was variable and did not always include the most antigenically distant viruses.



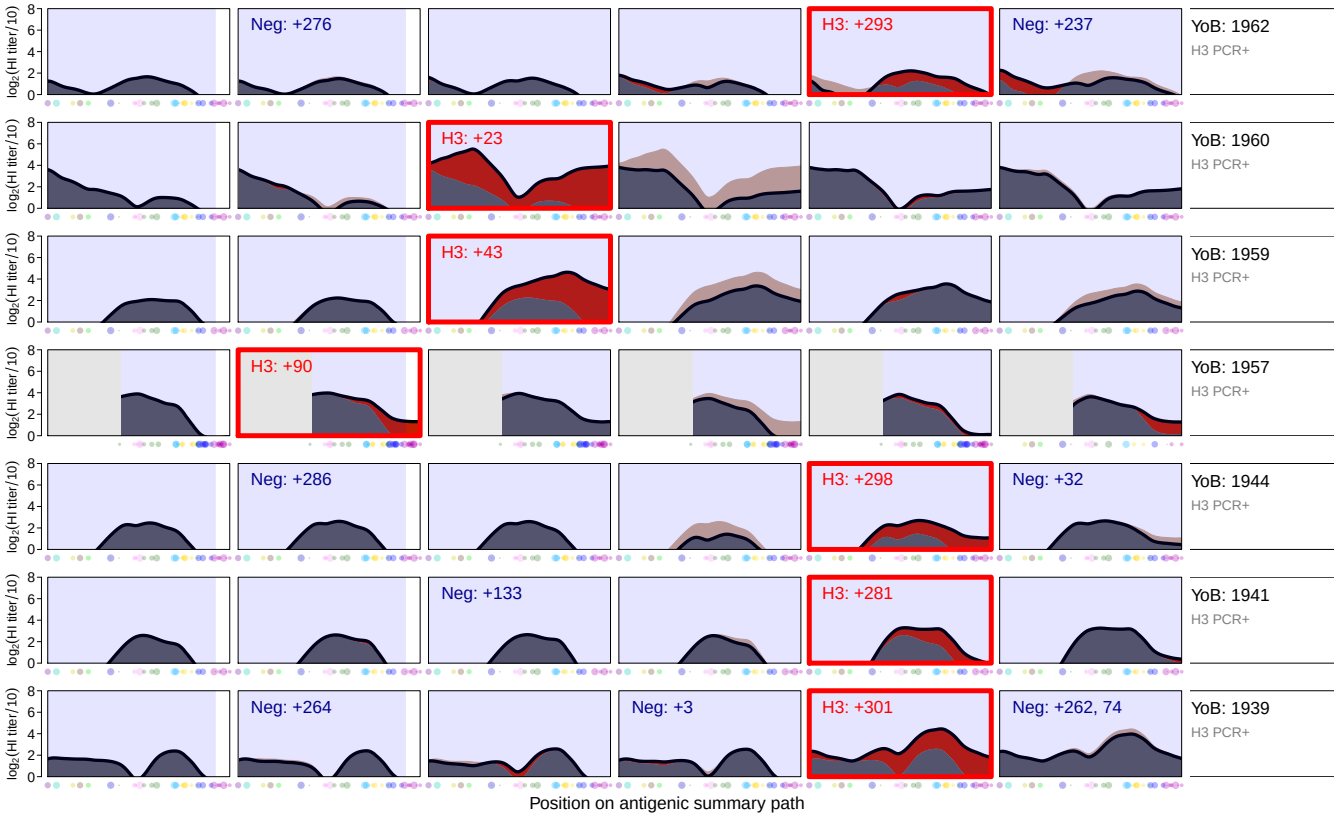



Figure S21: (see above, pages 28 to 29) Antibody landscapes for all individuals with PCR-confirmed A/H3N2 infection in the Ha Nam cohort (legend as Figure S15).

### 2.2.5 The antibody response in seroconverted individuals

In addition to PCR-confirmed cases of infection with A/H3N2 influenza, a number of individuals show marked landscape changes between years. The landscapes of individuals (n=36) that met the criterion for seroconversion (a four-fold or greater change in HI titer to the relevant reference virus titrated, see section 1.1.2) are shown in Figure S22. Some of these individuals were also PCR-positive for an associated symptomatic A/H3N2 infection, and are included in this analysis, but were also listed separately in Figure S21. However, 26 other individuals showed seroconversion without PCR-confirmed infection. These individuals did not report ILI, for example because their symptoms were too minor to consider reporting, due to non-compliancy in reporting, or because these were genuine asymptomatic infections.

Some individuals had reported ILI, but were PCR-negative for all influenza subtypes tested. This situation could result from the lack of an adequate swap taken in the right timeframe needed to detect infection for that individual. It may also have been that an asymptomatic or otherwise unreported A/H3N2 infection occurred at a different time between sample collections.

Most of the landscapes shown following seroconversion share the features of a broad back-boosting response following infection that can be clearly seen. One exception is the sample taken from an individual born in 1957 in 2008: although a four-fold change in the reference virus titer was present in this case, the majority of titers remain unchanged or reduced (the landscape shows signs of antibody decay). Conversely, there were two samples that follow an A/H3N2 PCR-confirmed infection, and both show visible landscape increases, but do not meet the criterion for seroconversion, and are therefore not included in Figure S21. This demonstrates how the classic criterion of evaluation of serology data of a four-fold change, based on a single virus, can lead to potentially misleading results.

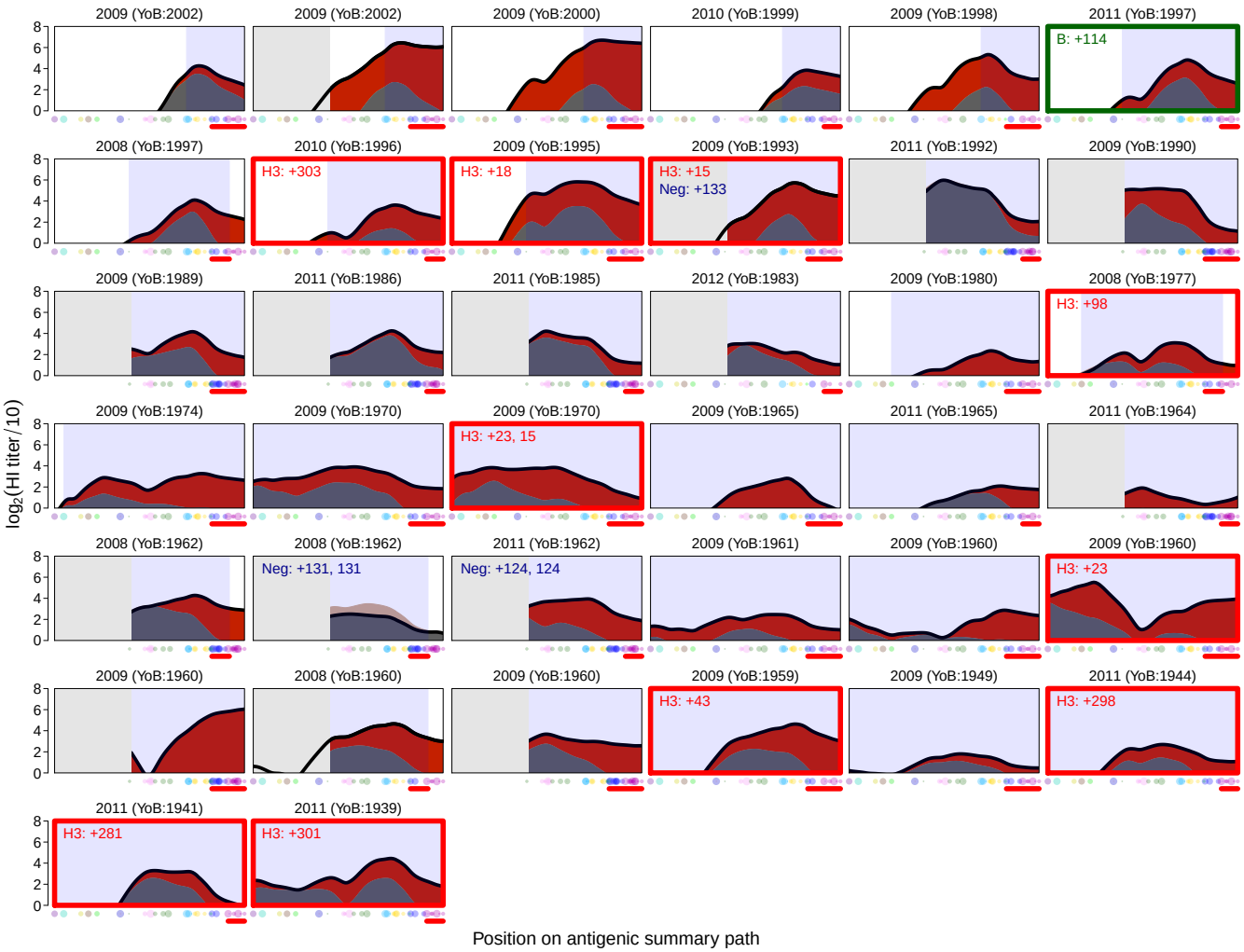



Figure S22: Antibody landscapes for all samples meeting the criterion for seroconversion (see section 1.1.2) of the Ha Nam cohort. Legend as Figure S15, except: the black line represents the landscape height along the path for the year given above each panel - this is the year of the sample from which seroconversion was detected. The red shading indicates increases, beige decreases, compared to the sample taken preceding seroconversion.

### 2.2.6 Antibody response following PCR-confirmed infection with other subtypes

Figure S23 shows the antibody landscapes for 2007-2012, where available, for individuals with any known infection of types and subtypes other than A/H3N2: PCR-confirmed A/H1N1, A/H1N1(pdm) or B influenza infection. The first sample after infection and number of days past ILI are indicated.

Examining Figure S23, most landscapes show no evidence of a response following infection with another subtype. However, the individual born in 1993 following an influenza A/H1N1(pdm) infection in 2011, and the individual born in 1977, 114 days following an influenza B infection in 2011, do show a change in antibody landscape. The finding that in many cases there are no changes seen post PCR-confirmed A/H1N1, A/H1N1(pdm) and influenza B infection is in contrast to the pattern following A/H3N2 PCR-confirmed infection, where changes were always apparent.

Given that none of the other individuals with PCR-confirmed influenza B, A/H1N1 and A/H1N1(pdm) infection show changes in antibody levels, and other studies in the same cohort also found an absence of cross-reactivity (*29*). Therefore, we would suggest that the individual born in 1993 with an influenza A/H1N1(pdm) infection in 2011, and

the individual born in 1977, 114 days following an influenza B infection in 2011, may also have been infected with A/H3N2 between sample collection dates, but were swabbed outside the time-frame necessary to detect the virus with PCR, or were infected at another date for which ILI symptoms were not reported. As discussed above, there have been other individuals that clearly seroconverted, yet were associated with PCR-negative results following ILI, or even in individuals without reported ILI.

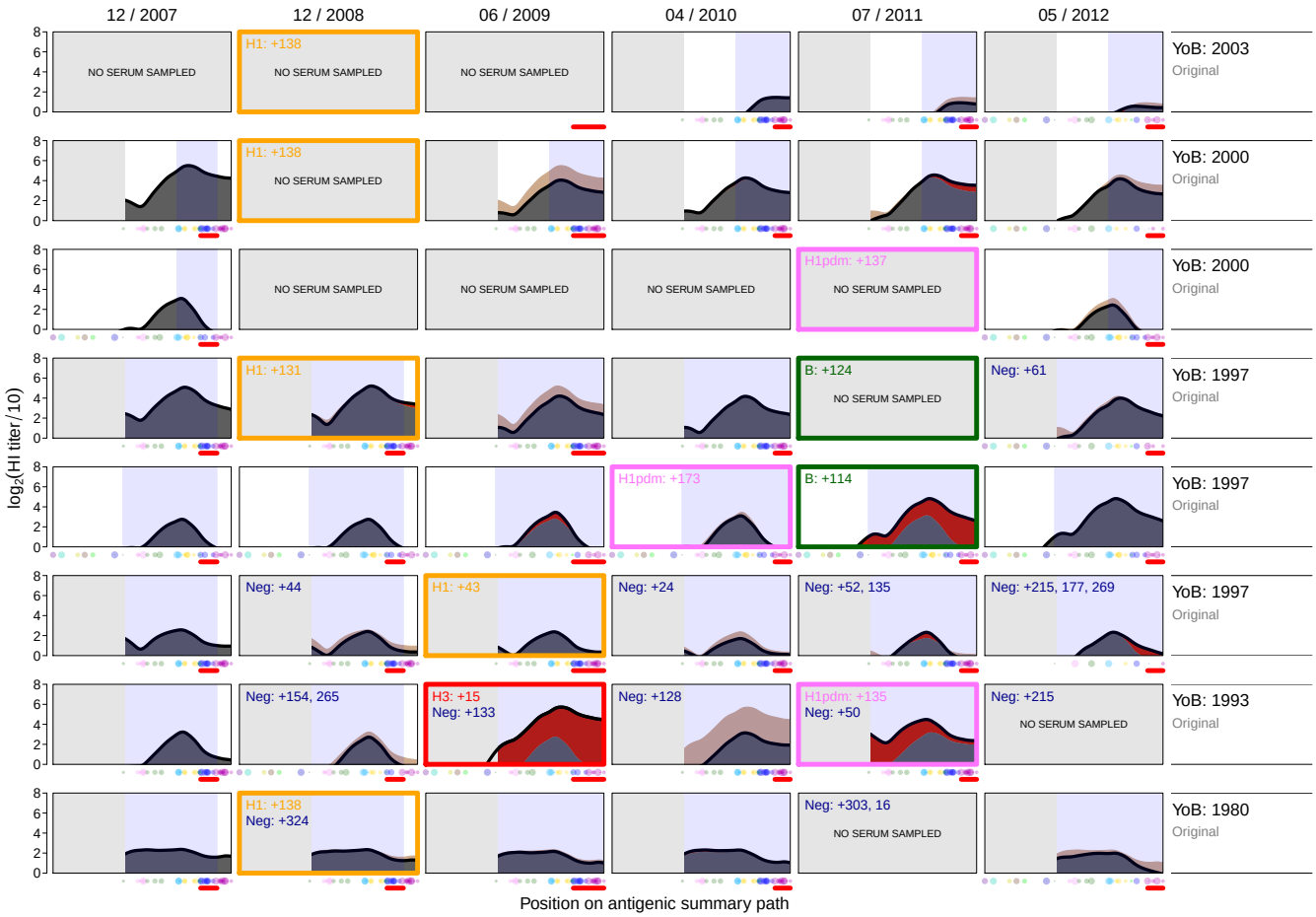



Figure S23: Antibody landscapes for all individuals with a PCR-confirmed influenza A/H1N1(pdm), A/H1N1 or B virus infection (legend as Figure S15). The first sample after a PCR-confirmed influenza virus infection is marked with a colored box, and the corresponding colored number gives the days after ILI: red corresponds to a sample PCR-positive for A/H3N2, orange for A/H1N1, pink for A/H1N1(pdm), green for influenza B. Samples that were PCR negative for any influenza subtype are indicated with a blue box.

### 2.2.7 Longevity of the antibody response: landscapes over time

To examine change in landscapes in relation to the time elapsed since the onset of symptoms, we used the information on the exact ILI date for individuals with PCR-confirmed A/H3N2 infection. Figure S24 shows the landscapes for these individuals, where the position of the subplot along the x-axis marks the time elapsed since the onset of ILI symptoms. Where available, subsequent samples are shown compared to the previous sample on the timeline to observe how a given landscape later decays over time. Individuals were ordered vertically based on the time elapsed between the onset of ILI symptoms and the time of sample collection. Note that landscapes from the same individual may be repeated in subsequent rows - their first sample following infection may be compared to pre-infection in one row first, then later, their second sample compared to pre-infection in its relevant "days since ILI" position. This repeated display creates a visualization of the timeline of changes as the time post-ILI increases, and maximizes use

of all the samples as there are multiple post-ILI landscapes per individual, corresponding to different numbers of days post-infection.


Figure S24: Antibody landscapes are shown based on the time since ILI for individuals with PCR-confirmed A/H3N2 infection. The antibody landscape in the last time point before infection is shown on the left, the first panel in each of the right-hand series of landscapes compares the post-infection sample with the pre-infection landscape. The number of days post-ILI is given in the top left corner, and individuals were arranged according to the number of days since the ILI. Each subsequent box compares the next sample taken from that individual to the previous one in the same row, to evaluate decay since the first post-infection landscape. Individuals are repeated when, for example, the second sample taken following an infection is compared to the pre-infection landscape in its appropriate position in the timeline (colors are used to mark the same individual). The black line represents the landscape height for a given year along the path. The red shading indicates an increase compared to the previous sample shown to the left in the same row and beige a decrease. Landscapes were cropped based on the available titrations performed for that time point.

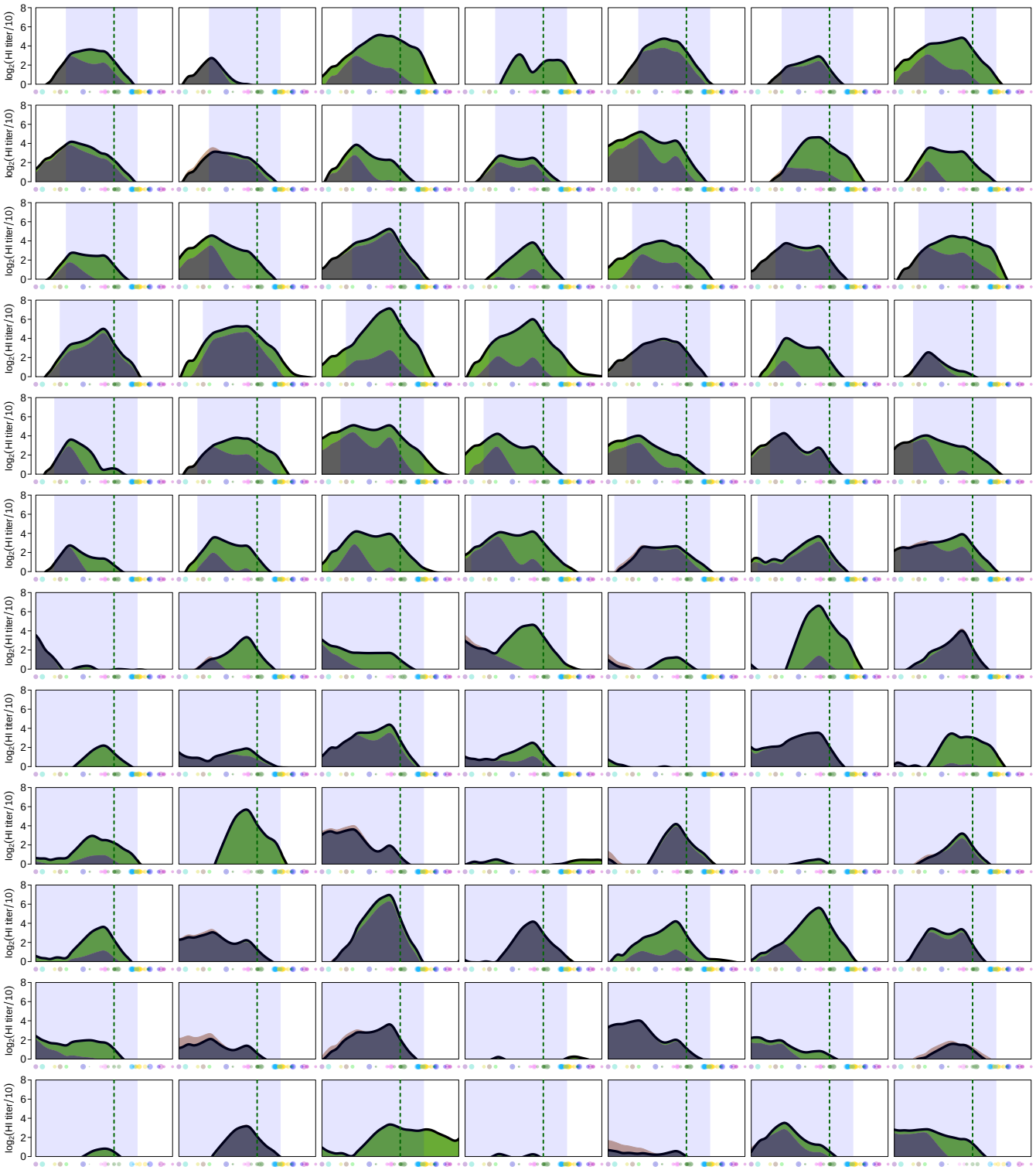
Figure S24 thus indicates how the magnitude and breadth of the response to infection depend on the time that has elapsed. Although the limited number of samples available for this analysis prohibits any firm conclusions, we observe several features of the time-decay. First, antibody responses were first detected between 1 and 17 days post-ILI, which is consistent with previous reports of antibody responses being detected 7 days post-vaccination (22). Second, initial antibody responses, i.e. landscapes for samples drawn soon after ILI onset, generally appear broader than the longer-term changes in antibody levels (also consistent with the analysis in Figure S45). Third, there is little evidence of further decay for samples beyond approximately a year post-infection, in line with the findings by

Horsfall *et al.* (*6*). Any post-infection landscape changes that are present at this point were typically maintained for the remainder of the study.

# 3 Response to vaccination

## 3.1 Plots of individual landscapes

The pre- and post-vaccination landscapes for each individual in the 1997 study, shown in Figure S25, can be used to investigate the response to vaccination with A/Nanchang/933/95.



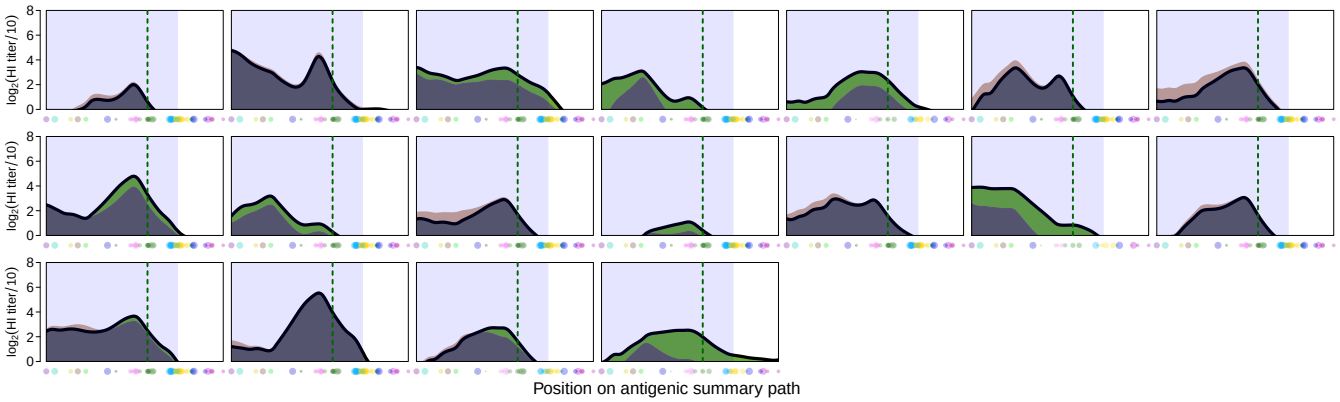
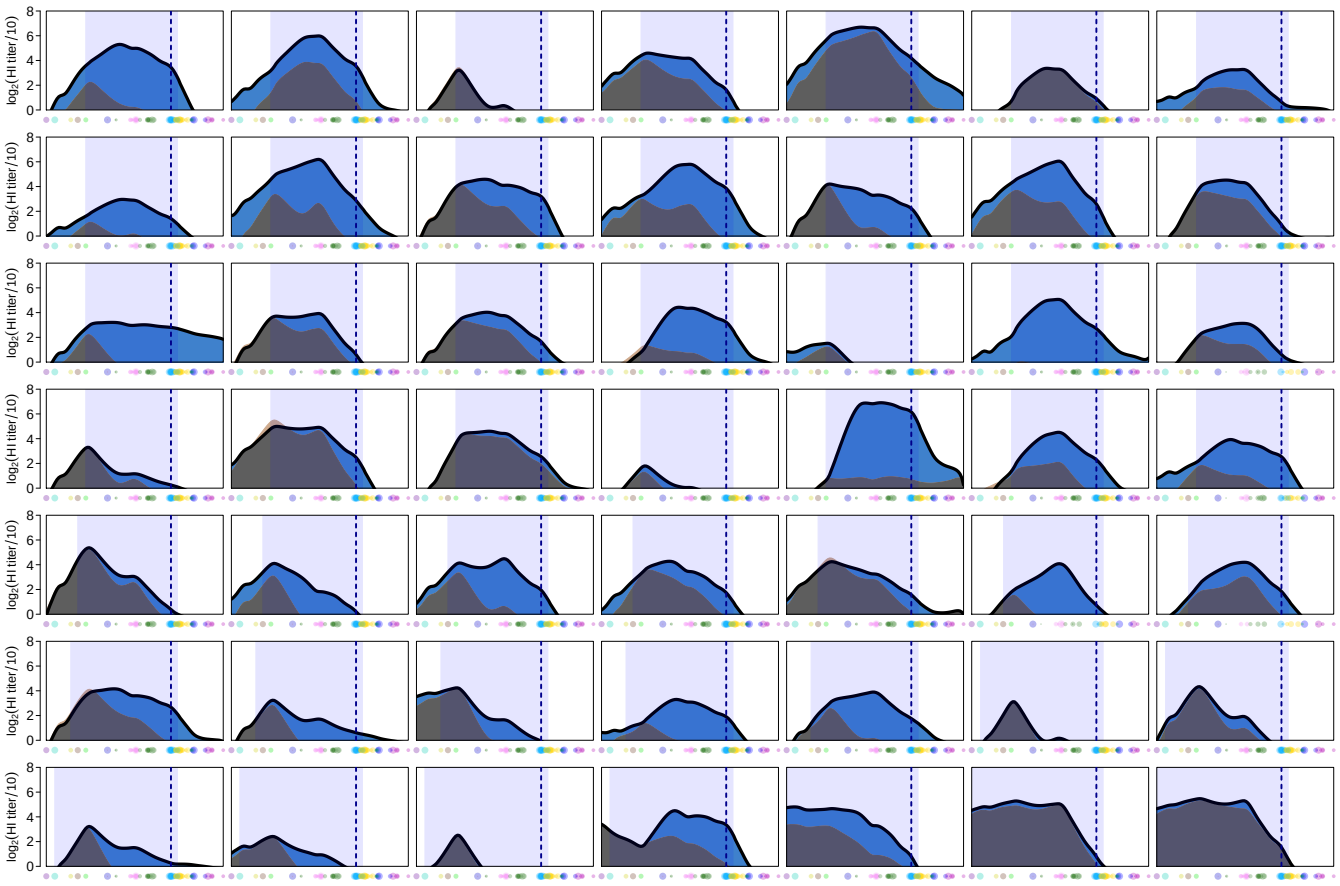



Figure S25: (see above, pages 34 to 35) Antibody landscapes for the 102 individuals in the 1997 study, vaccinated with A/Nanchang/933/95. The black line represents the pre-vaccination landscape height along the path; the green line represents the post-vaccination height. The green shading indicates an increase compared to the pre-vaccination sample and beige a decrease. The blue-shaded area indicates antigenic clusters that circulated during an individual's lifespan until sample collection. Titrated viruses are shown in the corresponding positions, where symbol size is scaled down with antigenic distance from the path. The dashed green line indicates the projection of the A/Nanchang/933/95 antigen location in the antigenic map on the summary path. The landscapes were ordered by increasing pre-vaccination antibody landscape levels to the vaccine strain. Note, unlike in the cohort study, each row of landscapes represents seven different individuals, not the same individual over time.

Figure S26 shows the results for individuals the 1998 study, vaccinated with A/Sydney/5/97.



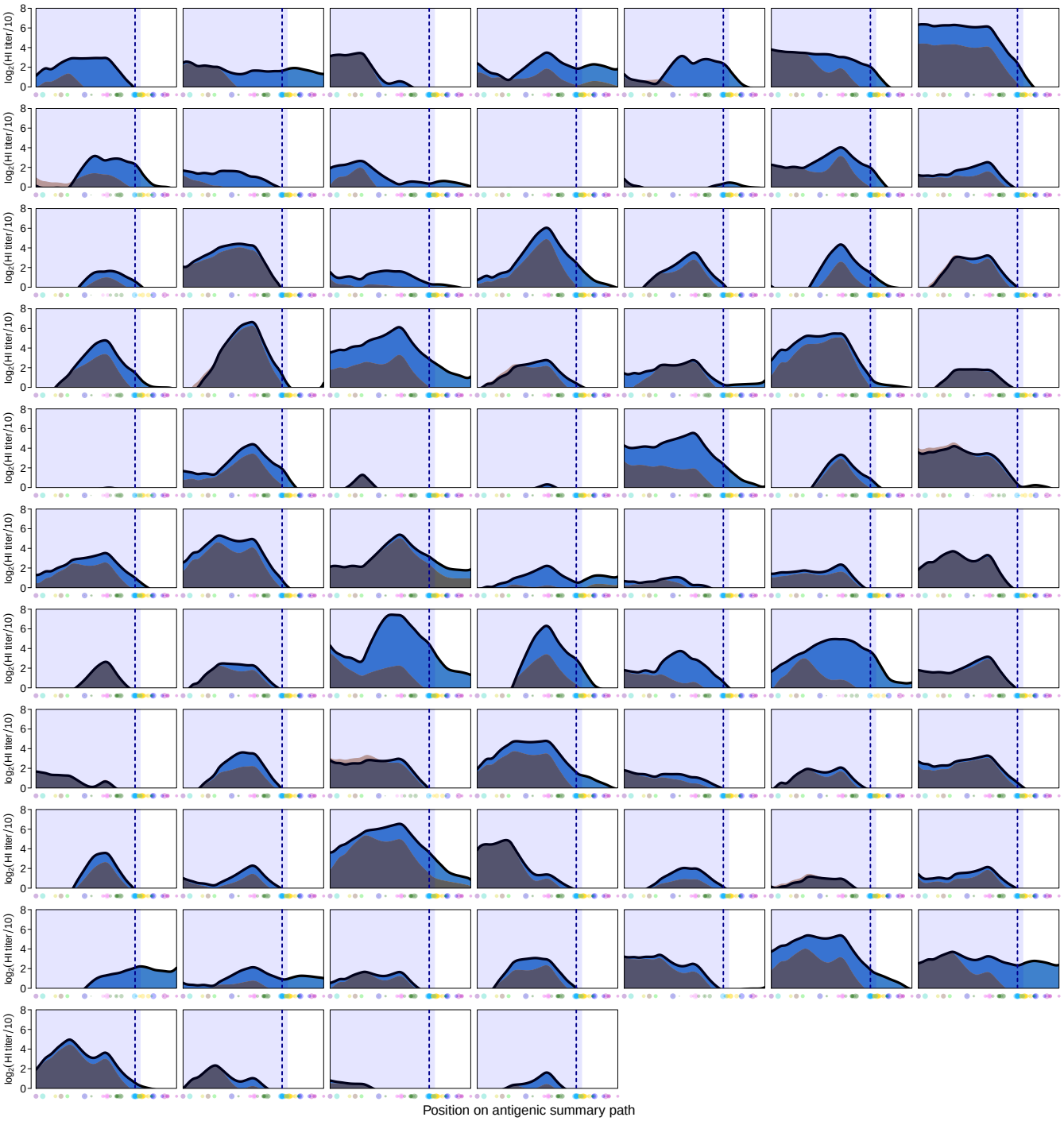



Figure S26: (see above, pages 35 to 36) Antibody landscapes for the 123 individuals in the 1998 study, vaccinated with A/Sydney/5/97. Legend as Figure S25, except: the black line represents the pre-vaccination landscape height along the path; the blue line represents the post-vaccination height. The dark blue shading indicates an increase compared to the pre-vaccination sample and beige a decrease. The dashed blue line indicates the projection of the A/Sydney/5/97 antigen location in the antigenic map on the summary path. The landscapes were ordered by increasing pre-vaccination antibody landscape levels to the vaccine strain. Note, unlike in the cohort study, each row of landscapes represents seven different individuals, not the same individual over time.

## 3.2 Analyses of vaccination data

### 3.2.1 Characteristics of the response to vaccination

As with infection, the response to vaccination in individuals showing a large titer increase for the vaccine antigen was strikingly antigenically broad, see Figure S25 and Figure S26. As in other vaccination studies, the magnitude of the response was highly variable among different individuals (Figure S27). In some individuals, there is very little or no change following vaccination, in others a substantial change is apparent.

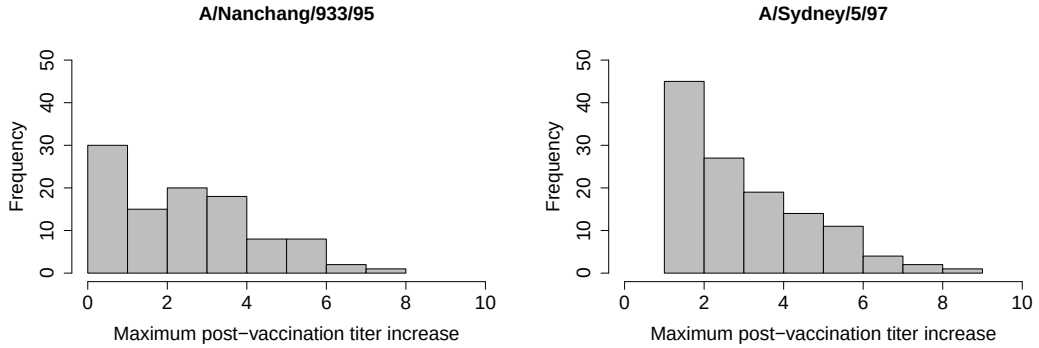



Figure S27: Histogram displaying the maximum titer fold-increase (based on HI titers) upon vaccination against any virus for the two vaccination studies.

The "antigen trapping" hypothesis, where the effective antigenic dose is lowered by binding of antigen by pre-existing antibodies and memory cells (*5, 7, 10, 19-20*), predicts a relation between the pre-vaccination antibody landscape, and the response observed in the antigenic region of the encountered strain. We also see evidence of a relationship between pre-vaccination titer and the increase in titers for the vaccine antigen upon vaccination: individuals with higher pre-vaccination titers typically displayed a smaller response to the vaccine (Figure S28).

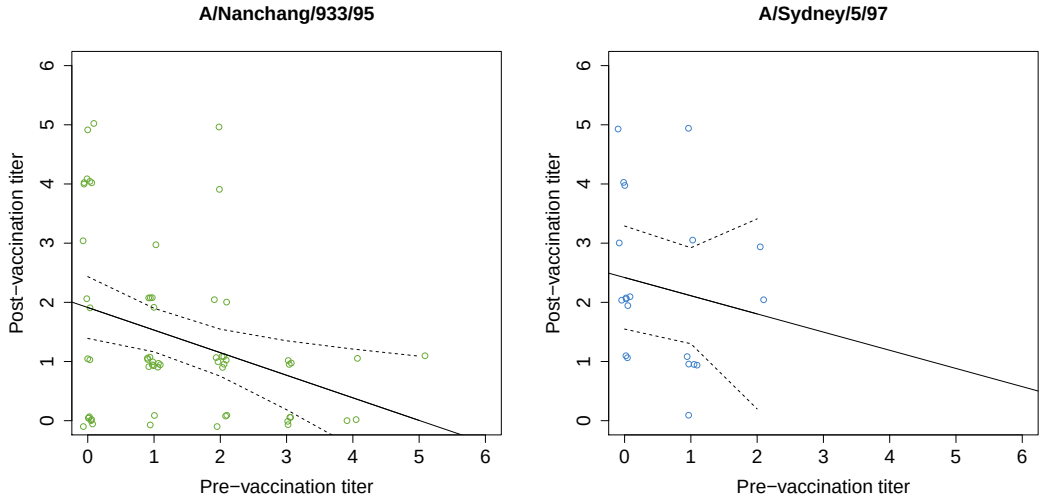



Figure S28: The pre-vaccination titer to the vaccine antigen is related to the response after vaccination (post-vaccination titer minus pre-vaccination titer) to the vaccine antigen, shown for the two vaccination studies (with jitter added to individual data points) by linear regression (solid line) and 95% confidence intervals (dotted lines).

### 3.2.2 Comparison of the response to infection versus vaccination

It is often presumed that the response to infection is more broad and/or strong than the response to vaccination. Our data sets were not directly designed to test this hypothesis, and a fair comparison of the antibody response to infection and vaccination is difficult because only symptomatic individuals were swabbed for PCR testing, and asymptomatic or mild cases may have gone unreported. Moreover, the vaccines contained a virus from the WU95 or SY97 clusters, whereas the infected individuals were exposed to viruses from the WI05 and PE09 clusters. Additionally, the precise virus in the vaccine is known, but although the range of likely infecting viruses can be estimated, the precise infecting virus is unknown. Finally, the time of sample draw after exposure was similar for all vaccinees (four weeks), but varied greatly for the infected individuals, where it would often exceed four weeks.

With these caveats in mind we nevertheless proceed to compare the response to vaccination and infection. We selected individuals with a substantial antibody response: we examined the responses of those individuals in the vaccination studies that seroconverted (a $\geq$4-fold titer change, i.e. $\geq$2 HI units, to the vaccine strain), and the response of individuals infected with A/H3N2 in the Ha Nam cohort study, established by either PCR-confirmation or seroconversion (change of $\geq$2 HI units to the relevant circulating reference strain, see section 1.1.2).

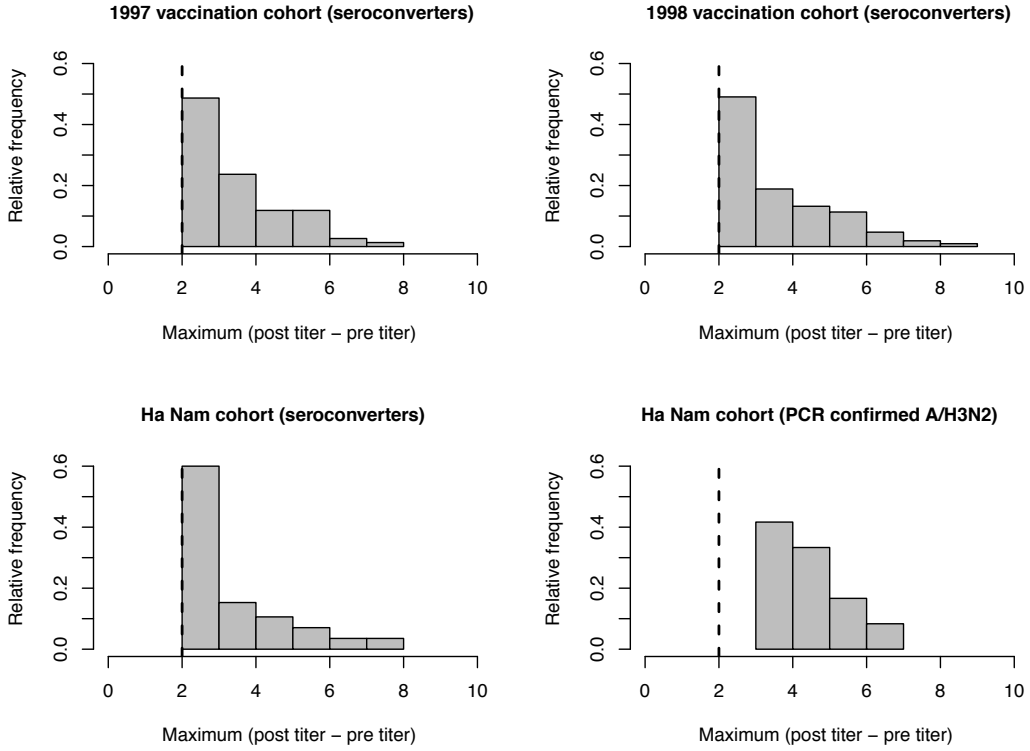



Figure S29: A comparison of the antibody response to vaccination and infection. For the two vaccination studies, individuals with a $\geq$4-fold change to the administered vaccine strain ($\geq$ 2 HI units) were selected, and the maximum response to any virus is given. For the Ha Nam cohort, the analysis was performed with individuals with a $\geq$4-fold change ($\geq$2 HI units, "seroconverters"), and for the subset of individuals with PCR-confirmed A/H3N2 infection. Note that these analyses give the maximum response, irrespective of antigenic location of this response.

Figure S29 shows that the distribution of maximum titer responses is skewed towards greater responses for individuals in the Ha Nam cohort following PCR-confirmed infection, as the density for post-titer - pre-titer differences that are 3 or higher (an 8-fold dilution change in the HI assay) is higher than in the other histograms of Figure S29. This is consistent with our previous observation that PCR-confirmed infected individuals often showed strong and broad antibody increases (except for the sample taken one day post-ILI). The three remaining plots all used the

same criterion to select “seroconverted” individuals, enabling a comparison of the vaccine and infection data sets. When comparing the plots of seroconverted individuals, it appears that the response to infection generally does not appear greater than the response to vaccination: the histograms look similar, although the infection landscape would typically be obtained longer post-exposure than the vaccination landscape.

## 3.3 Comparison of the two vaccination studies

### 3.3.1 Comparison of pre-vaccination antibody levels

The two vaccination studies consisted of two different groups of individuals, and samples were drawn one year apart (one study from 1997, the study from 1998). Thus, it was important to check for any difference in levels of pre-existing antibody prior to comparison of the responses to the two different vaccines. Figure S30 shows the difference between the mean pre-vaccination landscape heights of each study. In the region between the two vaccines (dashed lines), the pre-vaccination levels are very similar, and in the rest of the antigenic map the individuals in the 1998 (A/Sydney/5/97) study had minimally increased pre-vaccination landscape. The 95% and 99% confidence intervals of a t-test comparing the two different pre-vaccination landscape sets are also shown.

Given that higher pre-vaccination titers typically correspond to lower responses ((*27*), see also Figure S28 above), if there were any result from this small bias, it would be that the 1998 study, vaccinated with A/Sydney/5/97, should show a smaller response to vaccination, which is the opposite to what we observed. Thus, the increased response to A/Sydney/5/97 is unlikely to be an artifact of different pre-vaccination antibody levels.

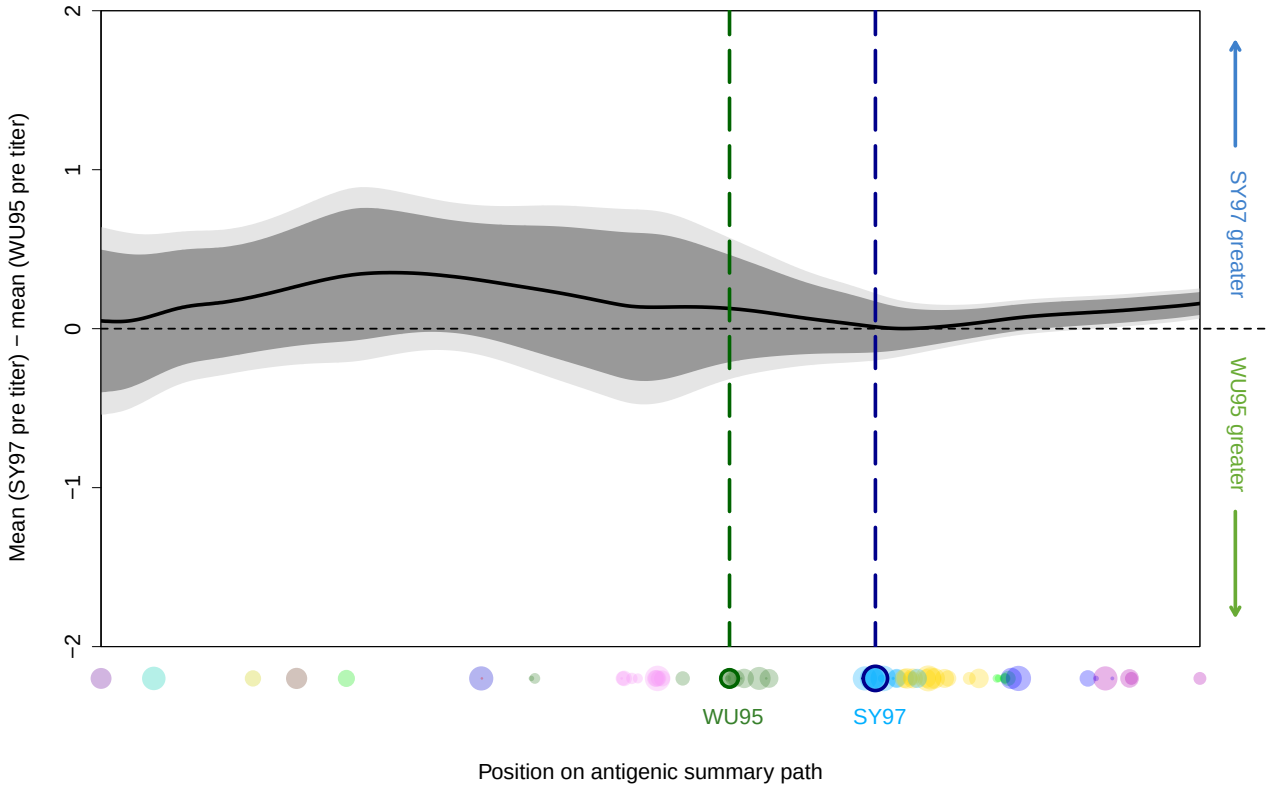



Figure S30: The black line represents the difference between the average pre-vaccination landscapes, for the 1997 (A/Nanchang/933/95) and 1998 (A/Sydney/5/97) studies. The 95% (dark gray) and 99% (light gray) confidence intervals of a t-test comparing the individuals in the two groups are shown.

To confirm that the pre-vaccination landscapes were indeed comparable, Figure S31 shows the HI values for each virus when compared with a two-tailed t-test.

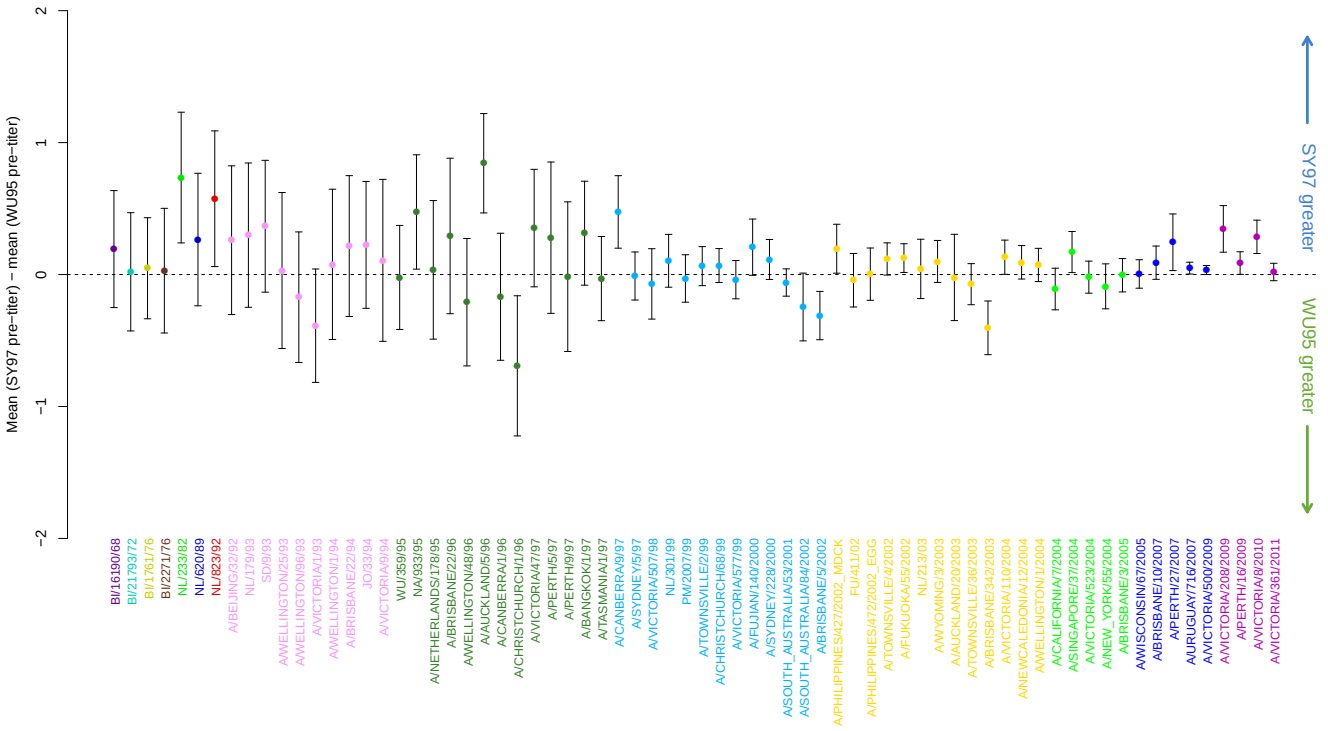



Figure S31: Comparison of the difference in HI titers between 1997 (A/Nanchang/933/95) and 1998 (A/Sydney/5/97) studies. The difference between the sample means of the 1998 and 1997 studies is shown as a point, colored by antigenic cluster. The bars indicate the 95% confidence interval of a two-tailed t-test for each antigen.

### 3.3.2 Comparison of antibody responses to the two vaccines

The comparison of the response to the two vaccination studies (Figure 3C, main manuscript) uses paired pre- and post-vaccination data: for each individual, the response to vaccination was calculated as post-vaccination minus pre-vaccination titer. Subsequently, the mean of these responses are compared for the two vaccination studies, by showing the difference between these two means: a positive titer difference indicates greater benefit from the A/Sydney/5/97 vaccination and negative indicates a greater benefit from the A/Nanchang/933/95 vaccination. The associated confidence intervals were based on t-tests comparing the responses of the individuals in the two studies.

In these calculations, it is difficult to deal with a point in the landscape where the individual had a non-detectable pre- or post-vaccination titer ($<$10), because such non-numeric titers are prohibitive for calculating the response (for a change from $<$10 to 20, all that can be said is that the response $>$2-fold). Rather than using a discrete arbitrary cut-off to select which data points to include when calculating the response, we weighted the values in the t-test based on pre-vaccination landscape height at each point as follows:

$$w_i = \begin{cases} 1 & \text{for} \quad z_i \geq 0 \\ z_i + 1 & \text{for} \quad -1 < z_i < 0 \\ 0 & \text{for} \quad z_i < 1 \end{cases}$$

where $w_i$ is the weighting for each point and $z_i$ is the associated height of the pre-vaccination landscape. Following this equation, if the pre-vaccination landscape height was $\geq 0$, the associated vaccination change is weighted 1.

If the pre-vaccination landscape was $\leq$ -1, this data point does not contribute to the calculation of the response to vaccination. Pre-vaccination titers between -1 and 0 are weighted following the linear relationship described.

We also repeated this analysis, by equating a pre-vaccination landscape value $\leq$ -1 to be -1. The results are shown in Figure S32, from which the same conclusions about antigenically advancing the vaccine are still drawn. This approach has the advantage of not biasing the responses measured to those with measurable (and therefore higher) pre-vaccination titers but necessarily loses power by assuming changes that occur below the sensitivity threshold of the HI assay are not present. The magnitude of difference in vaccination is therefore lower towards later antigenic clusters where titers were often below the detectable level and the confidence intervals are artificially narrowed. Despite these differences between the two methods of analysis, the same phenomena are observed, with an increased response to A/Sydney/5/97, importantly for the antigenic clusters circulating soon after WU95.

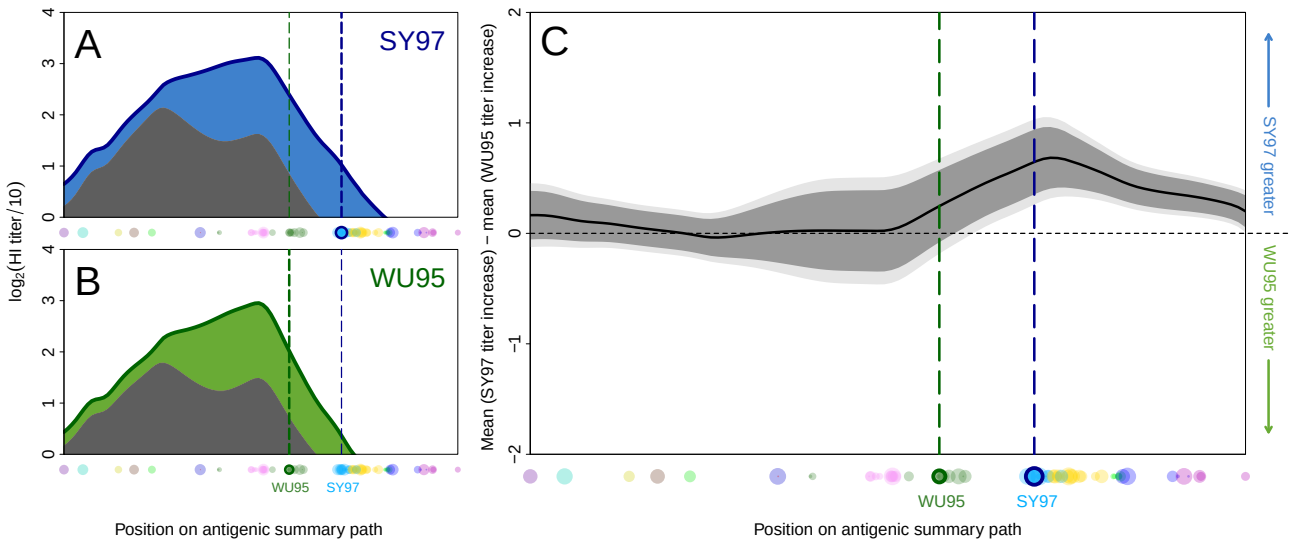



Figure S32: The average pre-vaccination landscape (black) and landscape after vaccination with (A) A/Sydney/5/97 (blue) in the 1998 study (123 individuals), or with (B) A/Nanchang/933/95 (green) in the 1997 study (102 individuals). (C) The average titer increase, i.e. the average difference between each individual's post- and pre-vaccination titers, following vaccination with A/Sydney/5/97 and A/Nanchang/933/95, with 95% (dark gray) and 99% (light gray) t-test based confidence intervals, where a predicted landscape height of $\leq$ -1 was treated as -1.

### 3.3.3 Comparison of the two vaccines across the full landscape

To assess whether the conclusions on antigenically advancing the vaccine held for all antigenic locations of the map, the difference in vaccination response was examined for each position of the landscape (not only along the summary path). Figure S33 shows this difference in color-code, and the absence of green regions, and prominent presence of blue regions clearly indicates that across all regions of antigenic space, responses were either equivalent or greater following vaccination with A/Sydney/5/97.

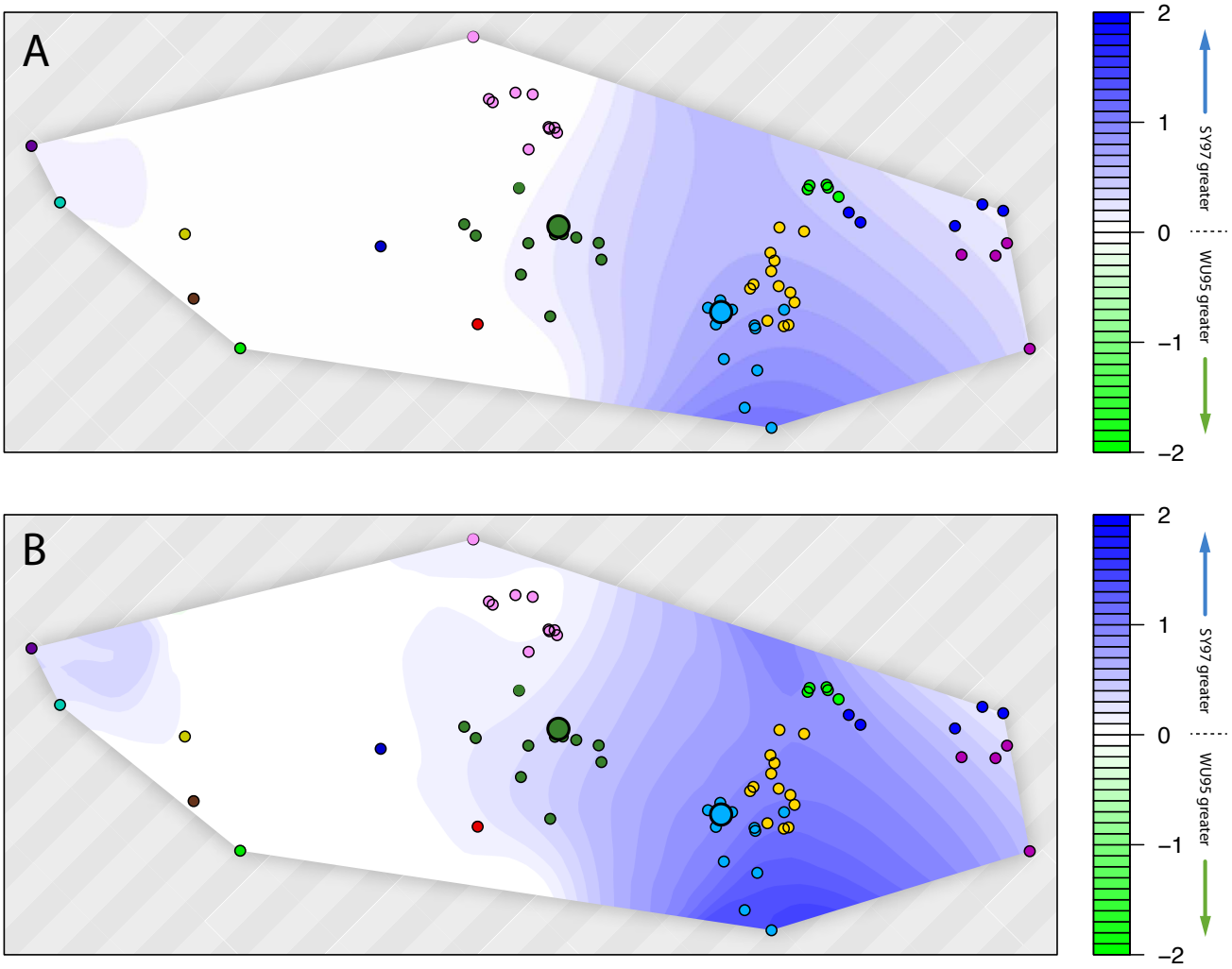



Figure S33: Antigenic map showing the antigenic position of viruses for which titers were determined in the vaccination studies, represented by dots color-coded by antigenic cluster; the two larger symbols indicate the two vaccine strains. The shading represents the difference between the mean increase in landscape height at that antigenic position for individuals vaccinated with A/Sydney/5/97 compared to those vaccinated with A/Nanchang/933/95 (scale given on right-hand side). Gray represents antigenic regions outside of the convex hull bounded by the viruses titrated, outside of which the antibody landscape would need to be obtained by extrapolation. (A) The result when treating landscape heights of <-1 as -1. (B) The result when weighting the values of landscapes below a detectable level, as described in section 3.3.2.

### 3.3.4 Comparison of the two vaccines based on the HI titers

The results found in the landscapes should also be found when examining differences in titer responses for the two different vaccines on a per-antigen basis. Figure S34 shows the beneficial effect of antigenically advancing the vaccine strain, in particular for the SY97 (cyan) and FU02 (yellow) antigenic clusters. As may be expected from the lack of landscape-based smoothing and no averaging out of measurement errors, the pattern is somewhat variable between antigens, however, the benefit of vaccination with A/Sydney/5/97 is still evident in the raw data. To correct for the multiple testing (t-tests were performed for 70 antigens), a Bonferroni correction was used to adjust the 95% confidence intervals:

$$\text{Overall confidence level} = 1 - \alpha \qquad \text{Individual confidence level} = 1 - \frac{\alpha}{m}$$

where $\alpha$ is 0.05 and $m$ is the number of confidence intervals to be calculated (m = 70).

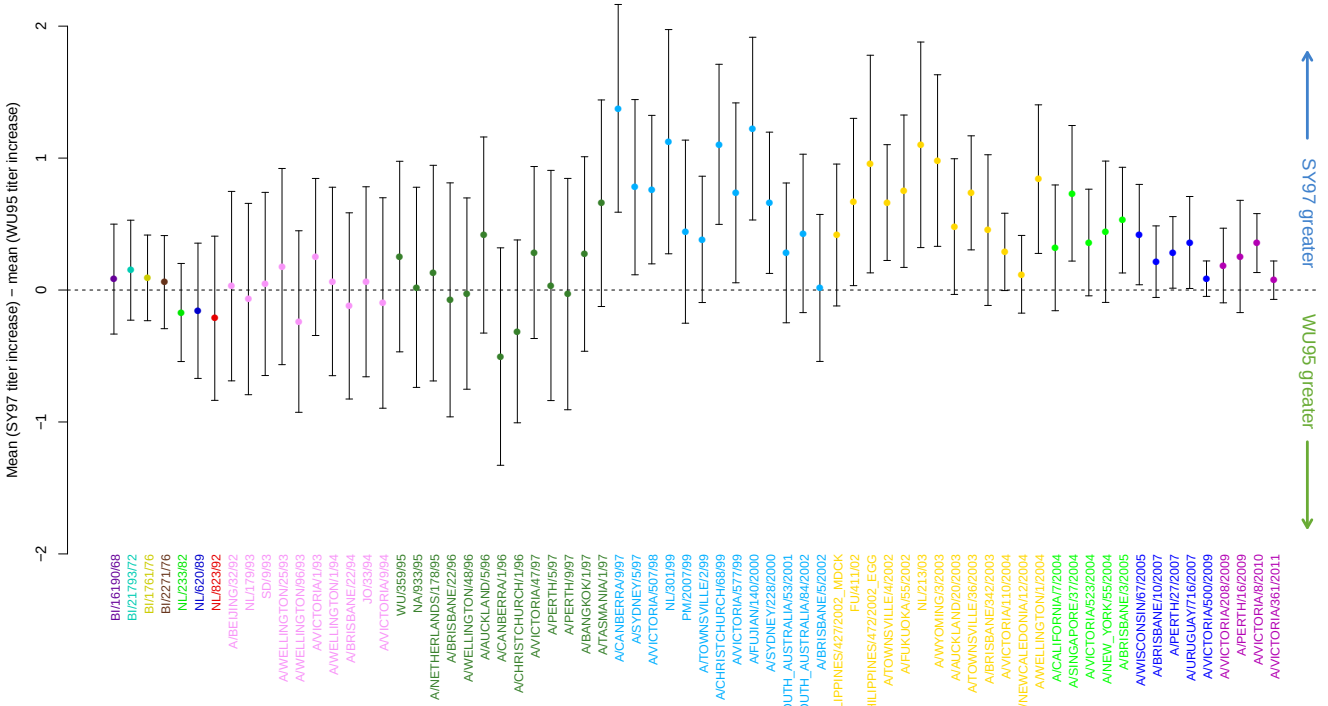



Figure S34: Comparison of the increase in HI titers between 1997 (A/Nanchang/933/95) and 1998 (A/Sydney/5/97) studies. The difference between the sample means of the response to vaccination in the 1998 and 1997 studies (based on the post-vaccination minus pre-vaccination titer for individuals in both studies) is shown as a point, colored by antigenic cluster. The bars indicate the 95% Bonferroni-corrected confidence interval of a two-tailed t-test for each antigen.

### 3.3.5 Comparison of changes in proportion of subjects with "protective" titers

An alternative approach to compare the responses to the two vaccines is to calculate the number of individuals with a landscape value corresponding to an HI titer of 40 or higher (which is associated with a 50% protection rate (*30*)). Figure S35 shows that the proportion of individuals with titers $\geq$40 is comparable pre-vaccination, and increases upon vaccination, where the proportion with such titers for the SY97 and FU02 clusters is markedly higher upon vaccination with A/Sydney/5/97. Even when considering only raw titers against the respective vaccine strains, there was no significant difference in the proportion of individuals with "protective" titers against the WU95 strain post-vaccination (67% in 1997, 71% in 1998, p-value=0.61) while the difference in proportion against the SY97 vaccine strain was marked (16% in 1997, 39% in 1998, p-value=0.0007). In each case the difference in proportion of "protected" individuals pre-vaccination was not significant.

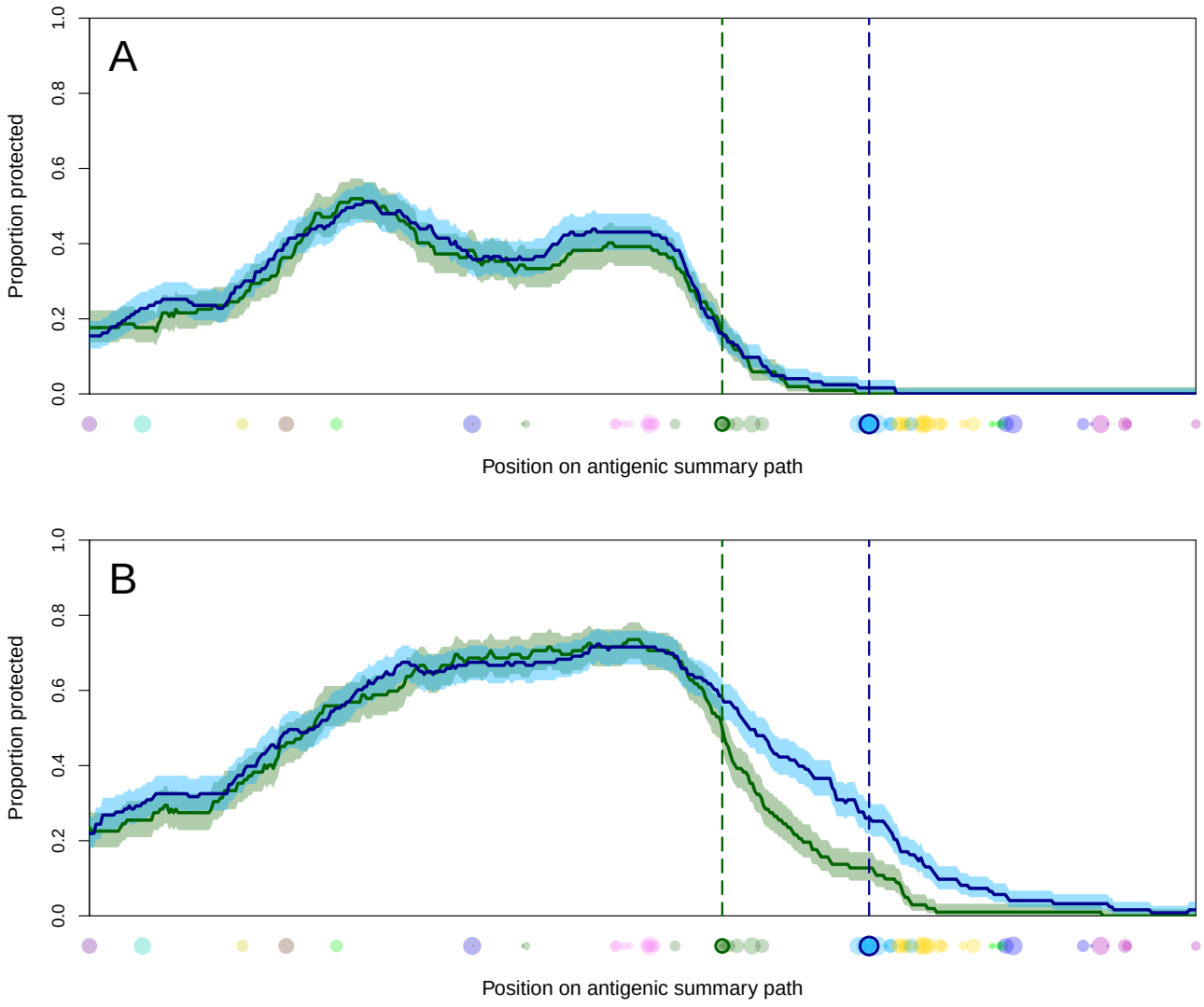



Figure S35: Comparison of the proportion of individuals with a landscape value $\geq$40 (2 HI units), for the studies on A/Nanchang/933/95 (green) and A/Sydney/5/97 (blue) in the (A) pre-vaccination landscapes, and (B) post-vaccination landscapes. Lighter shaded regions represent the estimated standard error of the proportion protected in each population at each position.

### 3.3.6 Analysis of responses to vaccination in elderly

To examine if the beneficial effects of antigenically advancing the vaccine strain are also observed for a group targeted in seasonal influenza vaccination campaigns, we repeated the analysis on the subset of individuals aged sixty or over. Figure S36 demonstrates that the findings of the full studies are mirrored when looking at the elderly only, and that, again, antigenically advancing the vaccine gives similar or improved antibody responses compared to the A/Nanchang/933/95 vaccine across the full evaluated antigenic range.

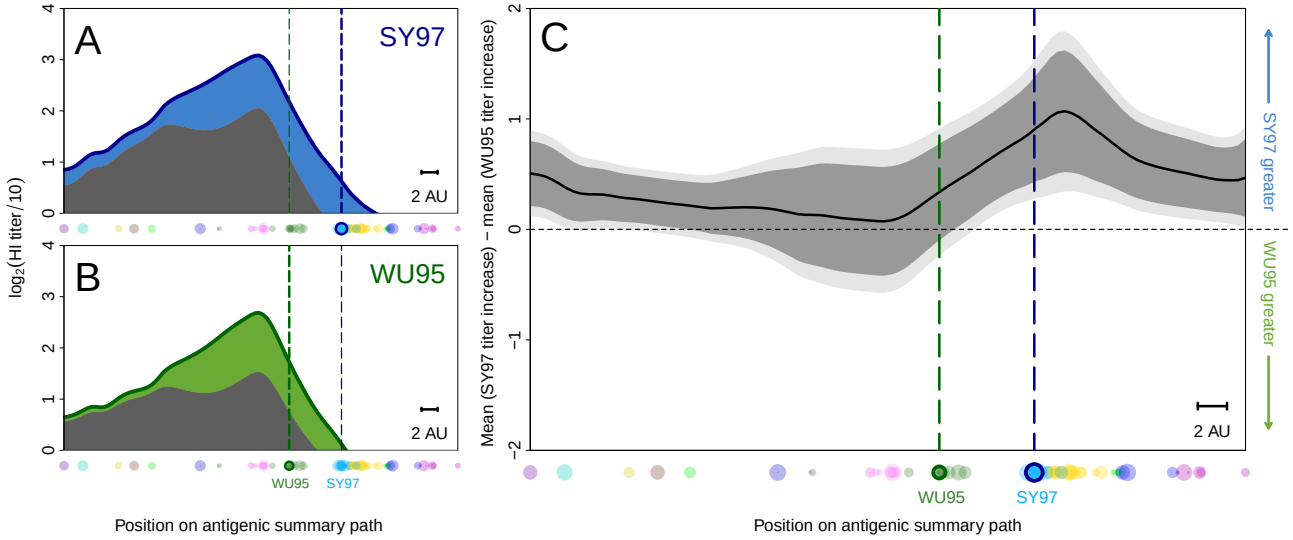



Figure S36: For individuals aged 60 or older, the average pre-vaccination landscape (black) and landscape after vaccination with (A) A/Sydney/5/97 (blue) in the 1998 study (62 individuals), or with (B) A/Nanchang/933/95 (green) in the 1997 study (57 individuals). (C) The average titer increase, i.e. the weighted average difference between each individual's post- and pre-vaccination titers, following vaccination with A/Sydney/5/97 and A/Nanchang/933/95, with 95% (dark gray) and 99% (light gray) weighted t-test based confidence intervals, where landscape heights of $\leq 0$ were weighted as described in section 3.3.2.

### 3.3.7 Analysis of responses to vaccination in at-risk individuals

Individuals with antibody titers $<40$ are traditionally considered to be most at risk from influenza infection. Thus, it is important to determine the antibody response specifically in this group, and to investigate whether the beneficial effect of antigenically advancing the vaccine strain is also observed for individuals with pre-vaccination titers $<40$ against A/Nanchang/933/95. Figure S37 demonstrates that this advantage is still observed.

Given that the back-boost appears to be dependent upon prior immunity, we also studied the responses for individuals with no measurable prior immunity, i.e. a titer $<10$, to the vaccine strain A/Nanchang/933/95. There were 41 individuals in the 1997 vaccination cohort and 38 individuals in the 1998 vaccination cohort with a $<10$ titer against the vaccine strain. Figure S38 shows the results for this subset of individuals (note that it differs from previous figures, as it also displays the landscape data below a the cutoff of HI titer 10, or 0 on the log-scale).

Although the benefit of the antigenically advanced vaccine is smaller for this subgroup, both groups still have a similar response against the WU95 cluster, and thus vaccinating with SY97 does not appear to compromise the response to WU95 even in this group of individuals with $<10$ pre-vaccination antibody titers. It could be that some low level of pre-vaccination antibody is still present, e.g. where landscapes are just below 0 on the log-scale, (and

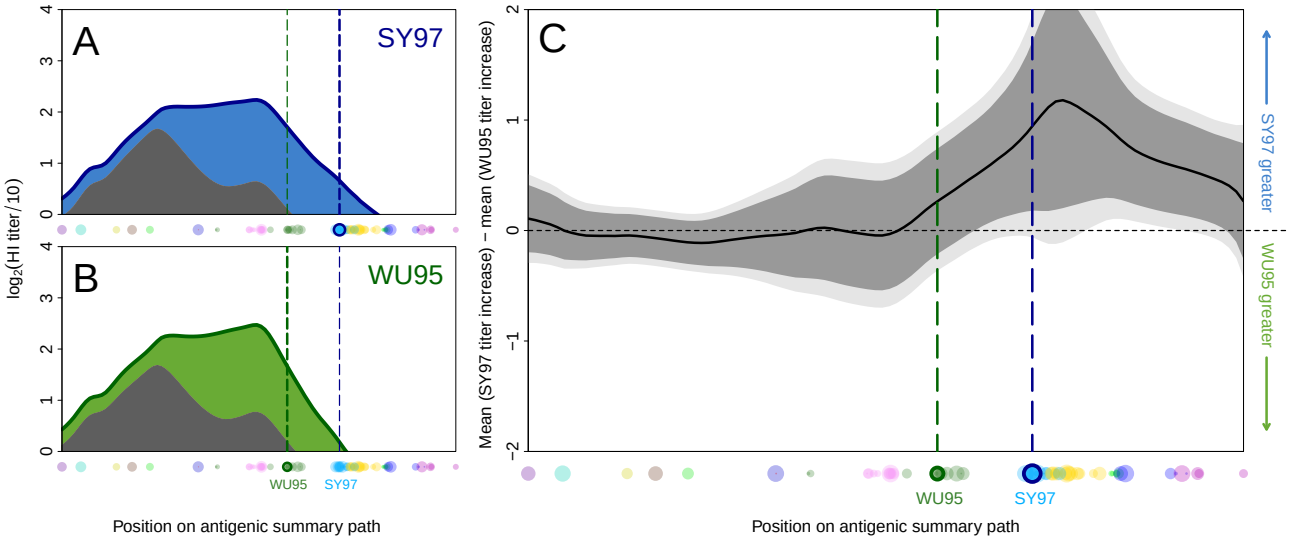



Figure S37: For individuals with A/Nanchang/933/95 pre-vaccination titers <40, the average pre-vaccination landscape (black) and landscape after vaccination with (A) A/Sydney/5/97 (blue) in the 1998 study (79 individuals), or with (B) A/Nanchang/933/95 (green) in the 1997 study (77 individuals). (C) The average titer increase, i.e. the weighted average difference between each individual's post- and pre-vaccination titers, following vaccination with A/Sydney/5/97 and A/Nanchang/933/95, with 95% (dark gray) and 99% (light gray) weighted t-test based confidence intervals, where landscape heights of $\leq 0$ were weighted as described in section 3.3.2.

titers can be e.g. 5 instead of 0, but both are measured as <10 in the HI assay). Interestingly, of the 79 individuals with a <10 HI titer against A/Nanchang/933/95, only 8 in the 1997 and 17 in the 1998 cohort have <10 HI titers for all of the WU95 viruses. The low, but present, antibodies that react with WU-95 like strains may therefore still lead to a small back-boost, and a net benefit of the antigenically advanced vaccine in this subgroup in terms of similar post-vaccination titers against WU95 and improved titers against SY97 with the SY97 vaccine, as compared to the WU95 vaccine, see Figure S38C.

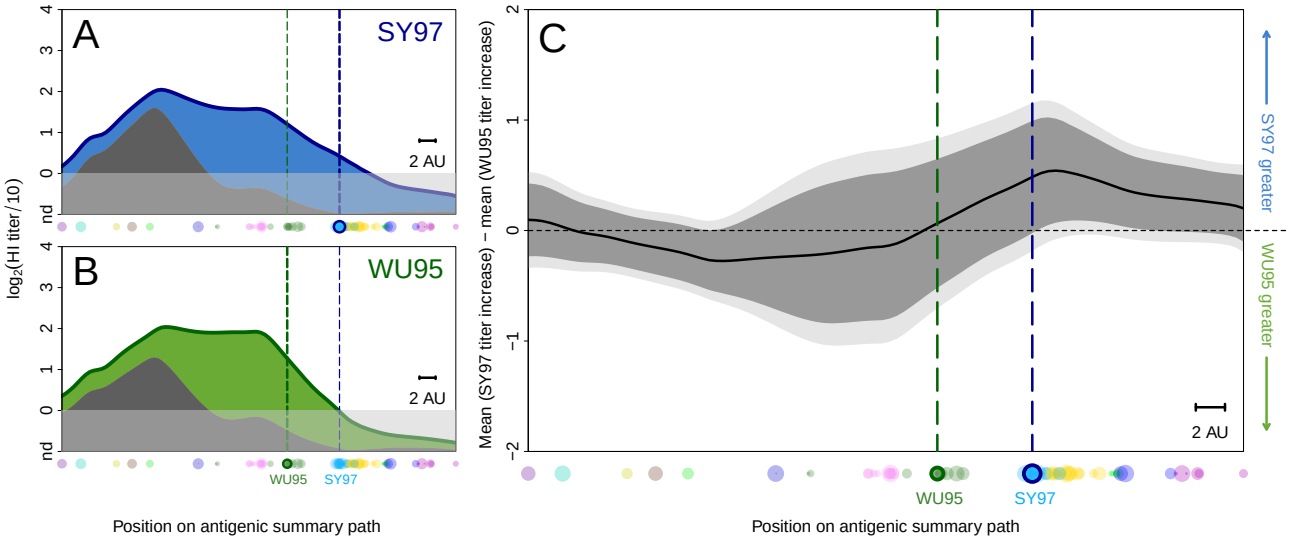



Figure S38: For individuals with A/Nanchang/933/95 pre-vaccination titers <10, the average pre-vaccination landscape (black) and landscape after vaccination with (A) A/Sydney/5/97 (blue) in the 1998 study (38 individuals), or with (B) A/Nanchang/933/95 (green) in the 1997 study (41 individuals). (C) The average titer increase, i.e. the average difference between each individual's post- and pre-vaccination titers, following vaccination with A/Sydney/5/97 and A/Nanchang/933/95, with 95% (dark gray) and 99% (light gray) t-test based confidence intervals, where landscape heights of $\leq$ -1 were treated as -1. The gray faded section of the plot in panels A and B represents the landscape below 0, a region where the landscape becomes somewhat unreliable as it contains more titers that are measured as <10, and thus the numeric value of the landscape cannot be accurately be inferred (nd: not detectable, i.e. <10).

### 3.3.8 Analysis of the 2009-2010 Wisconsin-Perth cluster transition

Using titrations from samples on two other annual seasonal influenza vaccine serology serum panels (samples from 2009 and 2010), we further investigated our findings in the context of another cluster transition (WI05 - PE09). Unlike the period between the 1997 and 1998 vaccination studies, in which the antigenically novel SY97 cluster had circulated at only very low levels when the 1998 study was done (*31*), there has been significant circulation of the PE09-like viruses between the dates of the two studies in 2009 and 2010 (*32*). This explains the increased levels of pre-vaccination antibody to PE09-like viruses in the 2010 cohort seen in Figure S39.

For several other reasons, the WU95-SY97 comparison described in the main manuscript is the better surrogate of evaluating the effect of antigenically advancing the vaccine: (i) the average pre-vaccination titer to the antigenically advanced SY97 strain was <10 in 1998, whereas as a result of the pre-vaccination antibody-reactivity from PE09 circulation prior to vaccination, the 2010 study less resembles the situation of "pre-emptive" vaccination; (ii) in the WU95-SY97 study, the pre-vaccination landscapes of the two groups were comparable, which is not the case for the WI05-PE09 study. This situation highlights the importance of a prospective study, where the two different vaccination approaches are tried in a clinical study by administering two groups different vaccines at the same time, to minimize differences in recent and previous exposure histories.

Retrospective examination of the WU95-SY97 and WI05-PE09 cluster transitions evidently show, as did the Ha Nam infection results, the back-boost stimulated by exposure (Figure S39). When investigating the pre-vaccination titer for the two different vaccine strains, the titer from the 2009 study against WI05 (dashed blue line) is comparable to the titer from the 2010 study against PE09 (dashed purple line). Thus, we would expect similar responses to the

two different vaccines, which is indeed observed, and is in contrast to the response to vaccination observed in the SY97 group in 1998, where the pre-vaccination titers against the vaccine were lower than the pre-vaccination titers in the 1997 group against the WU95 strains.

Because antibodies to the PE09 cluster had already appeared by 2010, Figure S39 represents what may be expected with the current vaccination strategy. From the results of the WU95-SY97 study, we think that if in 2009 the PE09 vaccine had been used, the overall antibody titers would have been higher. Conversely, we also expect that administration of WI05 in 2010 would have led to a response that is smaller than that resulting from the PE09 vaccine, because where vaccines lag behind influenza evolution, a loss in effectiveness is recorded. Given the higher pre-vaccination titers, and the negative relation between pre-vaccination titers and response to vaccination (S28), a smaller response to the same vaccine can be expected in 2010, than in 2009. Given the lower ability of response for higher pre-vaccination titers, when we consider the response after correction for the pre-vaccination titer, the "corrected" response is higher against the PE09 than the WU05 vaccine. However, the difficulties presented when comparing vaccinations in the context of differing pre-vaccination immunity complicate direct comparison, and again highlight the need for prospective clinical trials.

Figure S39: Comparison of two different vaccines spanning the WI05 - PE09 cluster transition. (A) The mean pre-vaccination landscape (gray) and landscape after vaccination with A/Perth/16/2009 (magenta, PE09 cluster) in the 2009 study (80 individuals), or (B) with A/Brisbane/10/2007 (dark blue, WI05 cluster) in the 2010 study (80 individuals). The 20 titrated virus strains are shown in their corresponding positions along the x-axis. The vertical dotted lines indicate the position of the WI05 (dark blue) and PE09 (magenta) wild type vaccine viruses. (C) The average titer increase between each individual's paired post-vaccination and pre-vaccination titers is shown with 95% (dark gray) and 99% (light gray) t-test based confidence intervals.

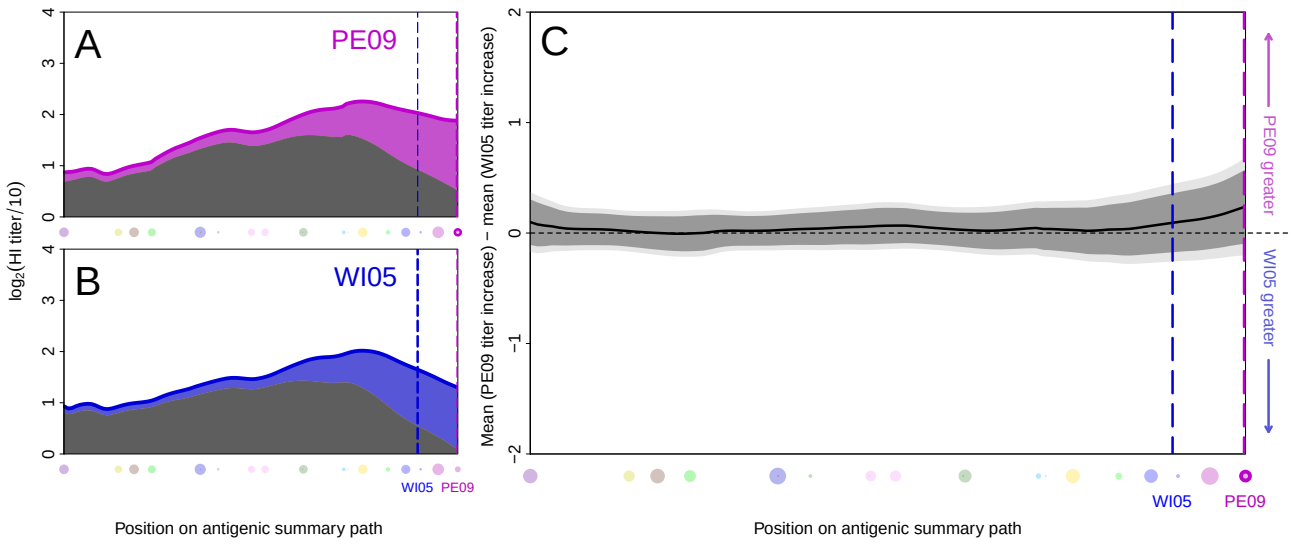



Figure S39: Comparison of two different vaccines spanning the WI05 - PE09 cluster transition. (A) The mean pre-vaccination landscape (gray) and landscape after vaccination with A/Perth/16/2009 (magenta, PE09 cluster) in the 2009 study (80 individuals), or (B) with A/Brisbane/10/2007 (dark blue, WI05 cluster) in the 2010 study (80 individuals). The 20 titrated virus strains are shown in their corresponding positions along the x-axis. The vertical dotted lines indicate the position of the WI05 (dark blue) and PE09 (magenta) wild type vaccine viruses. (C) The average titer increase between each individual's paired post-vaccination and pre-vaccination titers is shown with 95% (dark gray) and 99% (light gray) t-test based confidence intervals.

One suggestion for the differences between the responses to two WU95-SY97 vaccines is an intrinsic difference between the immunogenicity of the two vaccine strains. We deem this unlikely, because differences were antigen-specific and not present throughout the full antigenic range, and also because similar findings were made for the WI05-PE09 cluster transition study above. Moreover, Davenport *et al.* also found that an antigenically different vaccine may induce a greater response to the antigen of interest than an antigenically matched vaccine: they found that the A-prime vaccine induced lower antibody levels against A-prime strains than the antigenically unrelated "swine" vaccine strain (*9*).

### 3.3.9 Neutralisation assay results

To check the correspondence between findings made in HI and an alternative neutralisation assay, we performed virus neutralization tests against both vaccine strains (A/Nanchang/933/95 and A/Sydney/5/97) for pre- and post-vaccination sera from the 1997 and 1998 vaccination cohorts. Virus neutralization assays were performed by heating antisera for 30 minutes at 56°C. Twofold serial dilutions of the antisera starting at a 1:10 dilution were mixed 1:1 with 100 tissue culture infectious dose50 (TCID50) of the virus stocks. After incubation at 35°C for 2h in a 5% CO2 humidified incubator, the antiserum – virus mixture was transferred to 96-wells plates containing MDCK cells, which were washed twice with PBS prior to inoculation. Plates were incubated for 2h at 37°C, and inoculum was replaced by 200 $\mu$L infection medium. After four days end-point dilutions were read by hemagglutination assay.

Unfortunately, not all samples from the vaccine studies had sufficient volume to enable the virus neutralization tests, typically sera from high-responders that were previously used for other studies. This led to a sample selection bias, where there were relatively more low-responders than in the initial cohort, and was particularly present for the 1998 cohort. We were therefore able to test 80 of 102 sera from the 1997 cohort, and 69 of 123 from the 1998 cohort.

In our analyses, we found that both raw titers and vaccination responses as measured with HI and micro-neutralization were comparable (see Figure S40), giving an $R^2$ = 0.7698. Overall, it can be seen that virus neutralization titers tend to be higher than the HI titers.

We then compared the results of HI and neutralization for the pre- and post-vaccination titers per vaccination cohort, as shown in Figure S41, and noted that neutralization titers were typically higher, but the relative response patterns for the different groups remain similar. The mean post-vaccination titers against the WU95 vaccine strain A/Nanchang/933/95 were similar for the 1997 and 1998 cohort, whether measured by HI or virus neutralization. Thus, the antigenically advanced vaccine is still not at the detriment of the response to the previous antigenic cluster.

We noted in Figure S40 that a proportion of individuals who had measurable HI titers did not have titers when measured in the micro-neutralization assay. We therefore also wanted to check whether more of these lower responders as measured by virus neutralization titers (against WU95 strain A/Nanchang/933/95) were present in the 1998 cohort response than the 1997 cohort –because it is important to know if neutralization titers may not find consistently similar or greater responses to previous clusters with the antigenically advanced vaccine, a key finding to corroborate. To this end, we performed a proportion test comparing the proportion of subjects in each vaccination cohort that had a smaller response in the neutralization assay than in the HI assay, against the A/Nanchang/933/95 vaccine strain. Reassuringly, the proportion of individuals with lower neutralization than HI responses was not statistically significant being 24% in the 1997 cohort, and 28% in the 1998 cohort (p-value = 0.47).

In summary, the results show good correspondence between titers as measured by HI or micro-neutralization and give us no reason to doubt the conclusions regarding the benefits of vaccination with an antigenically advanced virus.

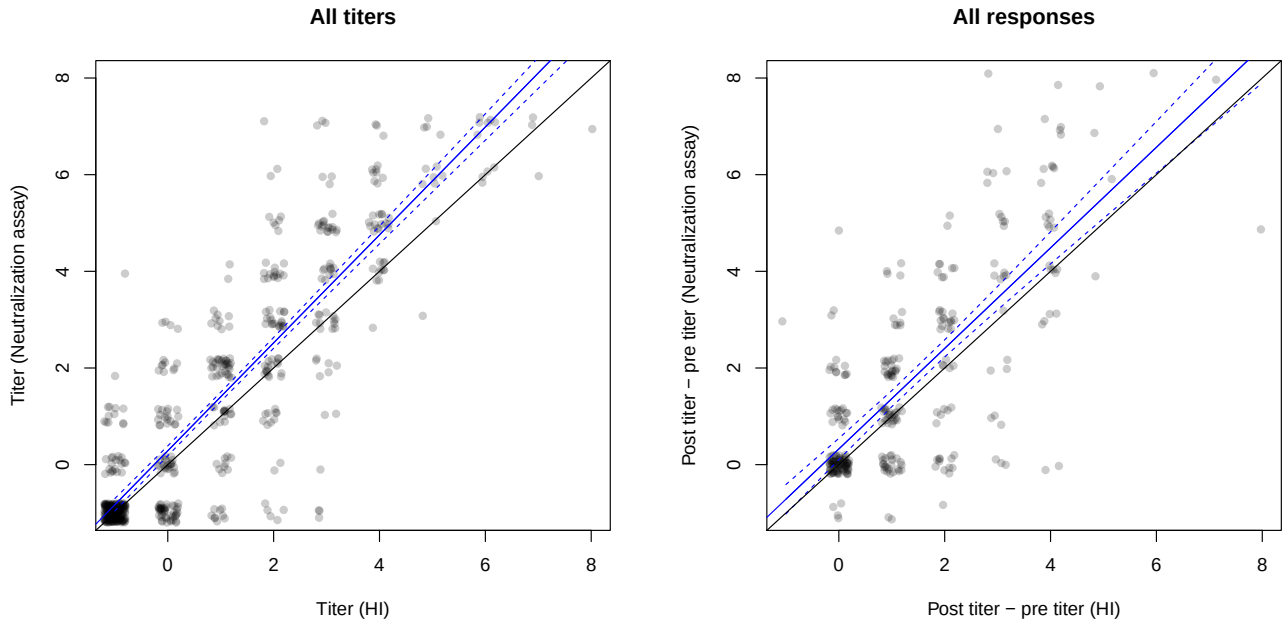



Figure S40: A regression analysis comparing HI titers with neutralization titers. Left panel: regression analysis of raw titers, plotting pre- and post-vaccination titers as measured by the two assays. Right panel: regression analysis comparing the vaccination response (post-vaccination - pre-vaccination titers) as measured by HI and virus neutralization assay.

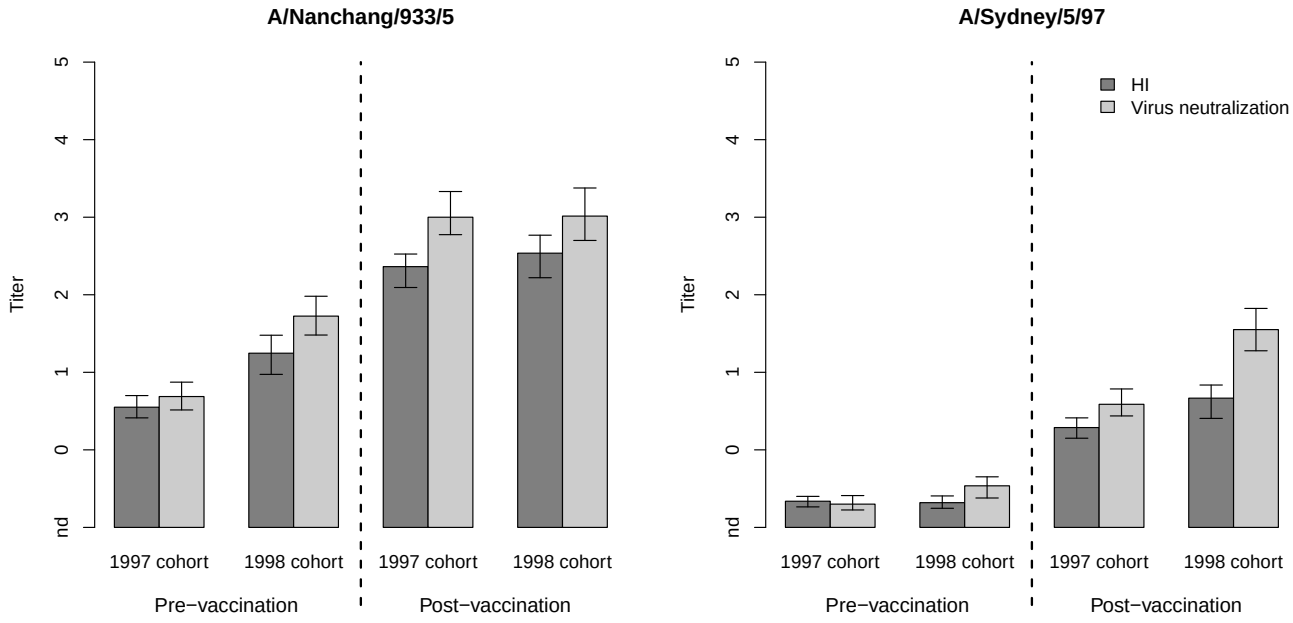



Figure S41: Mean pre- and post-vaccination titers in the subset of the 1997 and 1998 vaccination cohorts measured using both HI and viral neutralization. Error bars show the standard error based on calculations of the mean from 500 bootstrap samples from the each dataset with replacement.

# 4 Immunological and evolutionary interpretation

## 4.1 Comparison with "antigenic seniority"

Previous studies have also investigated patterns of influenza antibody reactivity within a population, for example Lessler *et al.* , who introduced the concept of "antigenic seniority" (*5*). Here the authors analyzed neutralization data from a range of serum samples as part of a cross-sectional study directed at an understanding of the patterns in the antibody reactivity present in the population against different influenza strains, but not specifically directed at post-infection responses. They found that in general, the titer to a given strain was lower as the age of the individual during the circulation of a strain increased. By plotting the pre-vaccination titers for the 1997 and 1998 vaccination cohorts and the samples taken in 2007 as part of the Hanam cohort in a similar fashion, we were able to compare our findings with those of Lessler *et al.* Figure S42 shows the HI titer for each virus compared to the age at the time of virus isolation for each individual.

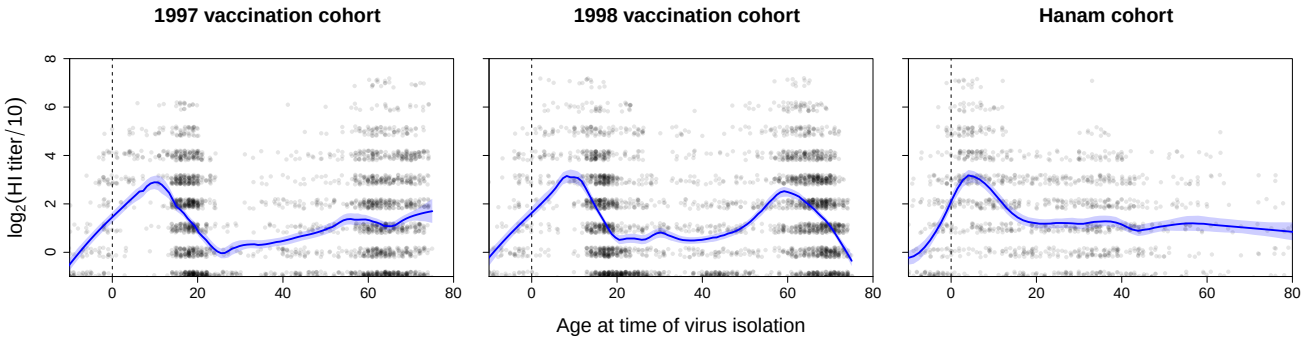



Figure S42: Plot of the 1997 and 1998 pre-vaccination data, and the Ha Nam cohort data in 2007 to enable comparison with the antigenic seniority study by Lessler *et al.* The plots show titers (with jitter) for each individual for each virus, as a function of the age at the time of virus isolation. A loess fit was applied to the data (span 0.25, R function "loess") and titers against strains that were isolated only after the year of sample collection were excluded.

From Figure S42, it is clear that titers isolated before an individual was born decline, as the gap between virus circulation and birth increases. We additionally calculated the maximum titer for each data set, and found this peaks at the age of 9.8 years for the 1997 vaccination cohort, 8.2 years for 1998 vaccination cohort and 4.1 years for the Ha Nam cohort. These are in agreement with the peak found by Lessler *et al.* at around 7 years of age. The earlier peak in the Ha Nam cohort may reflect different epidemiological patterns and first infections with influenza at an earlier age in this population. These data also indicate that the cross-reactivity reaches back by about 10 years on average, again similar to results by Lessler *et al.*

Interestingly, in the 1998 vaccination cohort, a second peak is visible around an age of 60 at virus isolation, and somewhat but not as clearly in the 1997 cohort, whilst is it not at all observed in the Ha Nam cohort. The absence of the peak in the Ha Nam cohort could be related to the absence of regular vaccination in Vietnam. These findings may suggest that the uptake of influenza vaccination in the other populations at around this age may be responsible for higher titers against the strains circulating in these age groups, as a result of seasonal vaccination.

We then examined the peak in each individual's landscape of these three cohorts in more detail. Rather than simply taking the average peak of the fit as we did above, we investigated the age of each individual during the circulation of viruses in the region of the peak landscape titers in Figure S43. The mode of this distribution was at age 3, and the median was at age 6, corresponding very closely to patterns of influenza exposure in children that have been reported previously, whereby over 60% of children have experiences their first influenza A/H3N2 infection by the age of 3 (*13*).

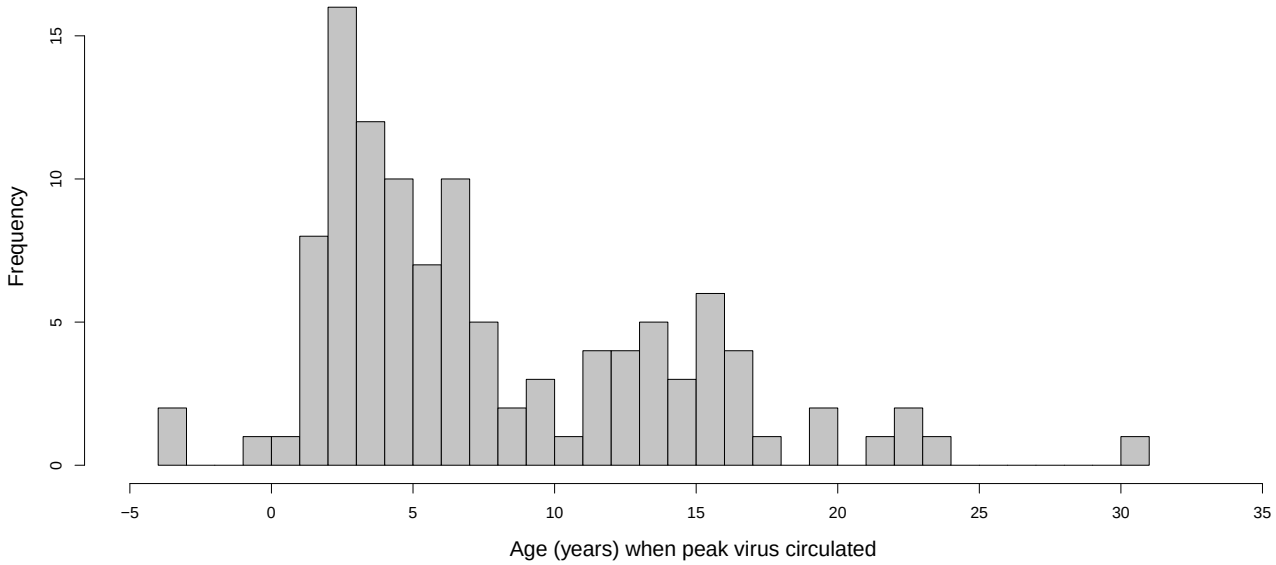



Figure S43: Age in individuals across all cohorts in the year of isolation of the influenza virus which corresponds to the highest region of their antibody landscape.

In general, these results support one interpretation of the term "original antigenic sin": the statement that a bias is present in the static amount of antibody towards an increased titer to the first encountered strain. In contrast, as we have already discussed extensively in the main manuscript, the antibody *response* to such strains upon influenza exposure is not biased to these first infection strains. In fact, the phenomenon demonstrated by Lessler *et al.* and in Figure S42 may be further evidence for the hypothesis of "antigen trapping", discussed above in relation to Figure S28. The antigen trapping mechanism would predict that the first infecting strain has generated more novel antibodies than subsequent exposures, since the first response would occur in the absence of pre-existing antibodies that have cross-reactivity and could "trap" antigen, and thus lower the amount of antigen available to stimulation a novel antibody response. Regardless of the underlying mechanisms, the lack of clarity when discussing static, absolute titers as measured in cross-sectional studies versus antibody responses to infection has created much contradiction and confusion, in particular in relation to the concept of original antigenic sin to date. We should therefore take care to separate these concepts clearly in future.

## 4.2 Mechanistic understanding of the back-boost

### 4.2.1 Comparison to primary immune responses

It is well established that a primary immune response to an antigenically variable pathogen produces a cross-reactive polyclonal antibody response, including even antibodies with heteroclytic properties that show improved binding to an antigenic variant other than the infecting strain that was used to create the primary response (*33*). This broad response is neatly illustrated in antibody landscapes of the first-infection ferret antisera (which were used to generate the antigenic map), as shown Figure S44. Cross-reactivity, as evaluated in terms of measurable titers (for ferrets infected with WI05 and PE09 strains), is seen for clusters such as FU02 and CA04, but not for antigenically more distant clusters such as BE92 and for the PE09-inoculated ferrets WU95.

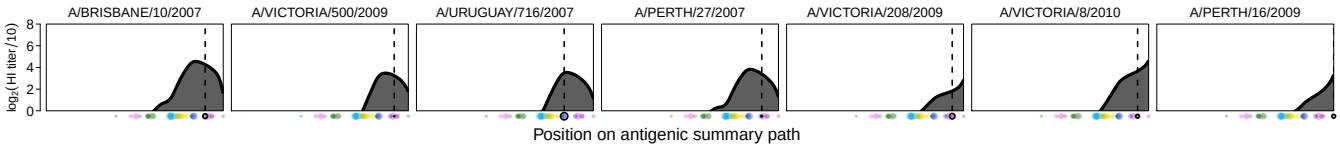



Figure S44: Antibody landscapes representing primary ferret responses to exposure to viruses from the Perth 2009 and Wisconsin 2005 antigenic clusters. The solid black line represents the post-inoculation landscape while the dashed vertical black line represents the location of the vaccine strain in relation to the summary path described in section 1.2.8.

Overall, detectable primary infection cross-reactivity in the ferrets tested extended to between 3 and 5 antigenic clusters before the cluster of the infecting strain. This is greater than the extent of detectable reactivity to clusters before an individual's birth as shown in Table S11, but this may be expected since ferret titers were measured at around the expected peak of the antibody response and it cannot be guaranteed that all of the human subjects would have even been exposed to the first cluster circulating after they were born. However, even at its peak, the extent of primary cross-reactivity is eclipsed by that of the back-boost, extending up to the full range of 14 antigenic clusters in some individuals.

### 4.2.2 Relationship between antibody response to infection and prior antibody levels

The extent of the response to infection with A/H3N2 was broad, but mostly limited to antigenic clusters that had detectable antibody levels prior to infection (HI titer $\geq$10). In addition to the visualization of this effect in Figure S21, we quantified the phenomenon in Figure S45, which shows the proportion of individuals with a detectable response, and the proportion of individuals with detectable pre-exposure antibody levels. Detectable titers were defined as a landscape value $>$0 (corresponding to HI titer $>$10), and a detectable response as an increase in landscape value $\geq$0.5 (where any landscape values $<$0 were treated as 0). Because these proportions decline similarly against older antigenic clusters, we hypothesize that the extent of the "back-boost" is determined by the extent of detectable pre-infection antibody reactivity. The same pattern is not observed for the more recent antigenic clusters (e.g. WI05, PE09), which could be the result of the generation of novel antibodies that are more specific to the infecting virus, rather than stimulation of prior immunity.

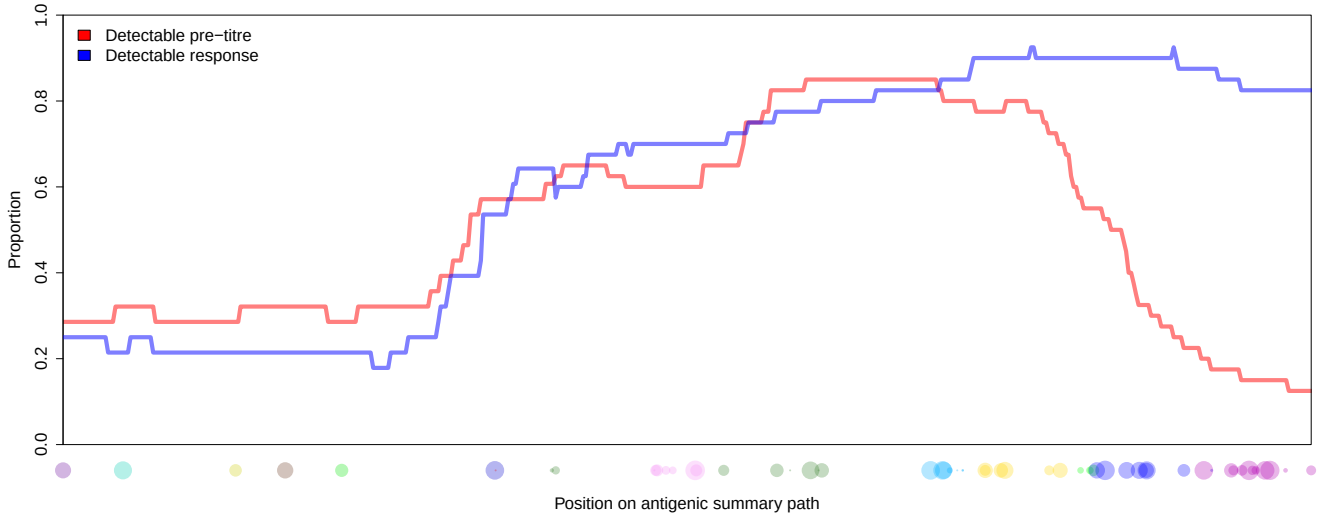



Figure S45: The proportion of individuals from the Ha Nam cohort with detectable titers prior to either PCR-confirmed infection with A/H3N2 or seroconversion (red line) is compared to the proportion of these individuals that show detectable increases in antibody landscapes (blue line).

### 4.2.3 Evaluation of different mechanistic hypotheses

The mechanism behind the broad back-boost described here is unknown, and the immunological nature of the response was not investigated as part of this study, however we considered a number of hypotheses and how well they were supported by the data.

First, we examined whether the extensive breadth of the response witnessed in the back-boost could be the result of the stimulation of naïve B-cells with extensive cross-reactivity, for example due to binding of the viral antigen to B-cells responsive to slowly evolving, recycled or highly conserved epitopes. We reasoned that this would predict responses that are not limited by the extent of pre-existing immunity, which contradicts our observations (see section 4.2.2), for example, we do not see responses against strains that circulated long before an individual's birth. The extensive cross-reactivity would instead give responses to any viruses that contain these slowly evolving, recycled, or highly conserved epitopes, even if they were not previously encountered. In addition, if the response were due to generation of broadly cross-reactive antibodies, one would expect an equal amount of titer increase across all antigens meaning that, on a log scale, the increase would be larger in regions of lower pre-infection titers (simplistically, because 10 pre-existing antibodies + 10 new antibodies is a two-fold increase, whereas 1000 pre-existing antibodies + 10 new antibodies is a 1.01-fold increase). However, in the back-boost responses, we observe homogenous responses (on the log-scale), rather than responses that are smaller (on the log-scale) when the pre-existing titers were higher. The exponential increases in HI titers across the regions of previously detectable HI titers therefore seem more indicative of a memory response, whereby memory cells already present are stimulated to undergo clonal expansion when infection or vaccination occurs.

If the broad back-boost is indeed part of a memory response, we reasoned this could be as a result of either a B-cell receptor binding independent mechanism, or as part of a response that is broad but nevertheless still requires some memory B-cell receptor binding affinity. A B-cell receptor binding independent mechanism could arise, for example, as a result of generalized stimulation of immunological memory cells as a consequence of inflammatory mediators associated with pathogen exposure. Such a mechanism would result in a broad increase of HI titers against the wide range of strains, and thus be consistent with the back-boost, and its limitation by pre-existing immunity. However, this mechanism would not explain the sub-type specific nature of the response. We therefore deem it more likely that the broad back-boost is the result of broad memory stimulation that requires at least some memory B-cell receptor binding affinity, i.e. a binding-dependent stimulation of memory cells. Naïve B-cell populations have been reported to require a higher binding affinity to be stimulated to clonal expansion than existing memory cells, and

thus are not likely part of the first response of the immune system upon a presenting antigen. Indeed, the broad back-boost is not at all displaying the characteristic shape that would be expected from a primary immune response, and that is seen in the ferret HI data that were used to generate the antigenic map. Differences in the binding affinity of B-cells for the encountered antigen also explain why antibody responses are extensive, but do not always include the most antigenically distant strains covered by an individual's immune repertoire.

### 4.2.4 Original antigenic sin and the back-boost

To characterize the breadth and longevity of the antibody response to infection on the background of prior immunity, the increase in antibody titer was determined for individuals that had measurable pre-infection titers. Figure S46 shows the titer increases for individuals with PCR-confirmed infection and for seroconverters, where individuals were excluded from the analysis at the point along the summary path where the predicted titers fell below 0 for the remaining antigenic region (hence differing sets of individuals along the summary path, and therefore the non-continuity in the plot). The responses of individuals with PCR-confirmed infection (after ILI, dashed lines) and those without PCR-confirmation, only seroconverting (dotted lines) are both included and follow a similar pattern.

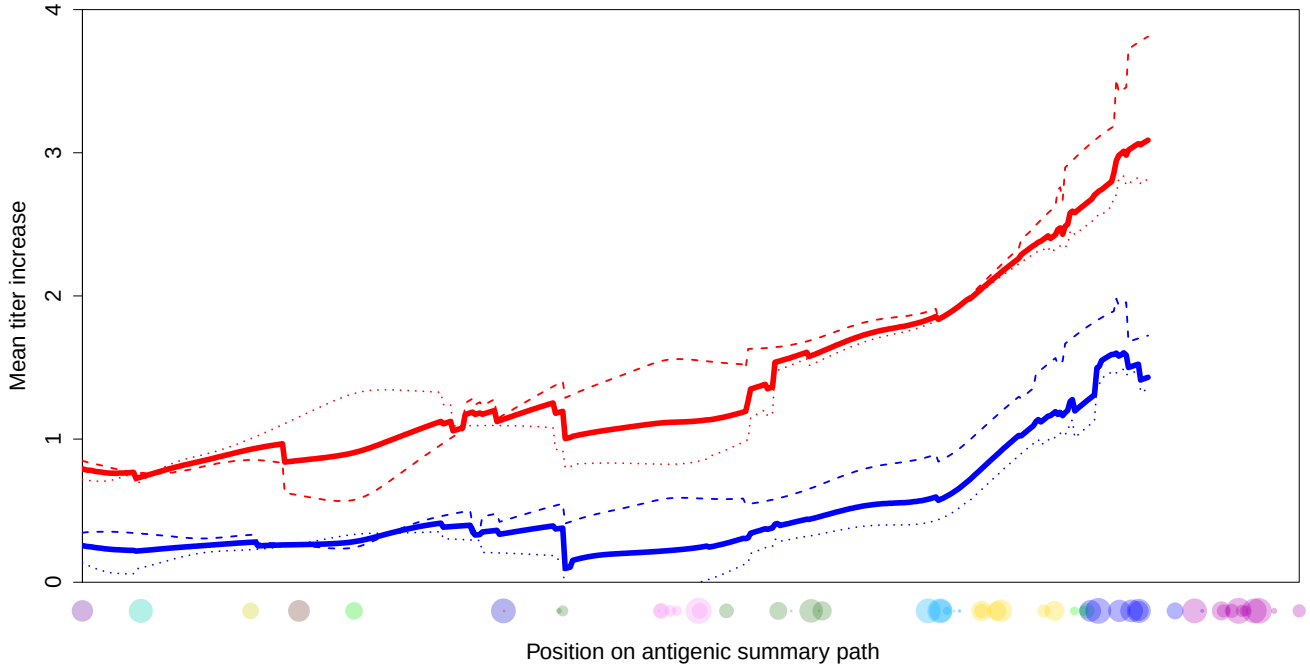



Figure S46: For each position along the summary path, the mean increase in titer was calculated based on a subset of individuals from the Ha Nam cohort that had a detectable titer at that point in the pre-infection landscape (i.e. pre-infection landscape titer was 10 or higher); based on PCR-confirmed infected individuals (dashed lines), seroconverters without PCR-confirmation (dotted lines), and the combination of these two subsets (solid lines). The red lines compare the sample before and immediately after infection, while the blue lines indicate the titer increase between the last available sample (often from 2012) and the antibody titers in the year prior to infection (when this was $\geq$1 year following the infection).

The initial response to infection (red line) is greater than the long-term increase in titers maintained beyond a year (blue line); there is a clear overall decay of the initial post-infection response (see also section 2.2.7). In many individuals this decrease reflects a return to pre-infection titers against older viruses over time, where the greater increased response to contemporary strains is maintained long-term as an antigenically narrower overall landscape increase. In some individuals however, increased titers even to the antigenically older regions are somewhat retained –for this reason, the average long-term responses approach but do not reach 0 in these earlier regions.

The magnitude of response typically declines as antigenic clusters become more dissimilar to the likely infecting virus, which in each case would be a virus from either the WI05 (blue) or PE09 (purple) antigenic clusters. Although the response is still substantial to older viruses, responses are largest for more recent viruses –in contrast to the prediction from the original antigenic sin hypothesis; while the back-boost is different from a typical primary immune response (see Figure S47).

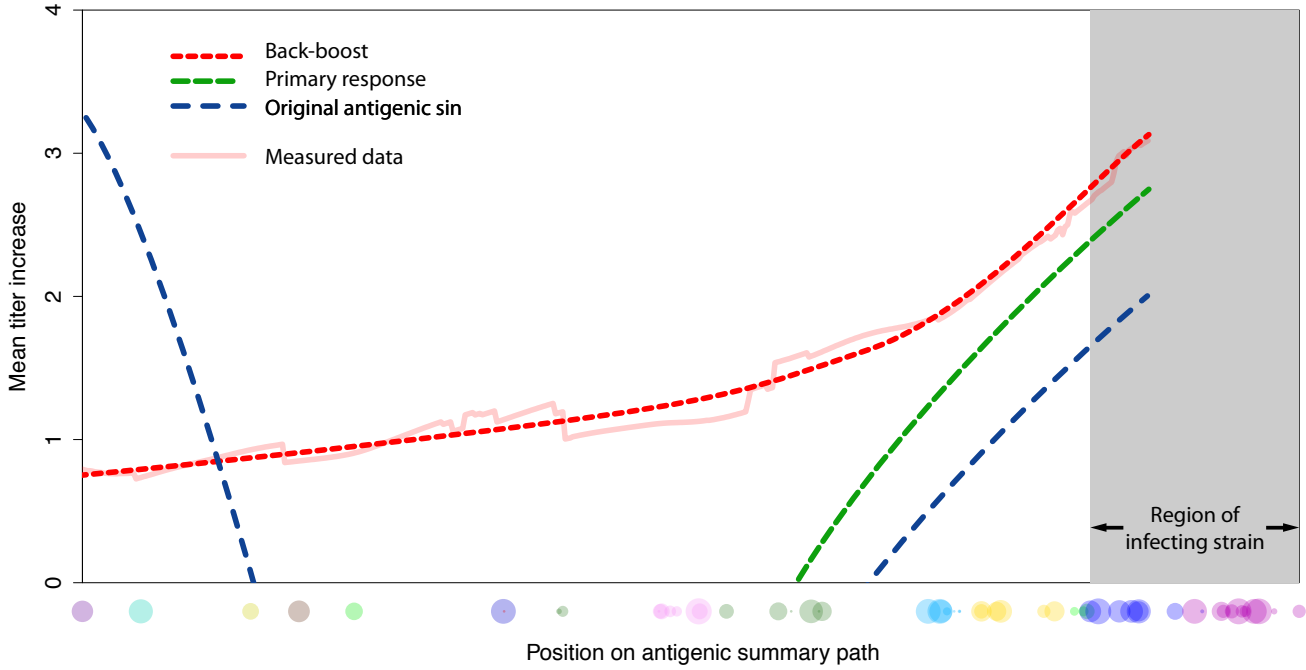



Figure S47: For each position along the summary path, the mean increase in titer was calculated based on a subset of individuals from the Ha Nam cohort that had a detectable titer at that point in the pre-infection landscape for individuals that were born before 1968, and thus alive during the HK68 pandemic, following PCR-confirmed A/H3N2 infection or seroconversion (faint red line). We then represent schematically the back-boost measured as measured in these responses (red line) and the prediction of titer increase according to the original antigenic sin hypothesis (blue line) and a typical primary immune response (green line).

### 4.2.5 Building-up of antibody landscapes over time

The static appearance of titers being highest against strains that circulated early in an individual's life has led to previous hypotheses that such patterns are the result of long-term and progressive reinforcement of antibody titers against earlier viruses upon exposure to each subsequent antigenic variant over time, see Figure S48A. If the back-boost were a long-term response, than this would indeed support this mechanism, as it would boost all antigenically senior strains. However, the back-boost changes over time, and instead we find that although the long-term increase in the antibody landscape can still span multiple clusters, it is more limited to the antigenic region of the likely infecting strain (e.g. Figure 2, rightmost column, and Figure S46). This provides evidence *against* the hypothesis of long-term and progressive reinforcement of the full breadth of antibody titers, and instead suggests that changes in serological immunity to strains that are antigenically more distant from the infecting virus are not maintained over time.

Combining our findings on the nature of the backboost with our findings in support of an "antigen trapping" hypothesis in relation to vaccination, we propose that the lifetime patterns can be alternatively explained by long-term increases which do not ultimately effect the response to very old strains, but where the strength of the response is

dependent on antigen trapping. Because the immune response to primary exposure is larger than the responses to subsequent exposures, due to complete absence of antigen trapping in the first exposure, one would indeed observe the highest static titers for the antigenic clusters encountered early in life. We do therefore agree with previous suggestions of "immunologic boosting and interference", since it is similar to the hypothesis that the antigenic trapping mechanism governs the strength of the long-term response (*5*). Figure S48B shows a schematic of this alternative and proposed building-up of the landscape over time, leading to the same static lifetime patterns as explained by panel A and as observed in previous studies (*5, 7-9, 11*). Note that they may still initially be a period of "reinforcement" of antibody titers whereby backwards cross-reactivity of antibodies produced against novel but antigenically similar viruses serves to increase overall serum antibody reactivity towards the first infecting strain - the extent of this period will depend upon both the rate of antigenic evolution and the degree of cross-reactivity. However, as antigenic distance increases, this effect ultimately diminishes leaving only transient changes in the titers for the first infecting strain upon later infections, as we observe in this study.

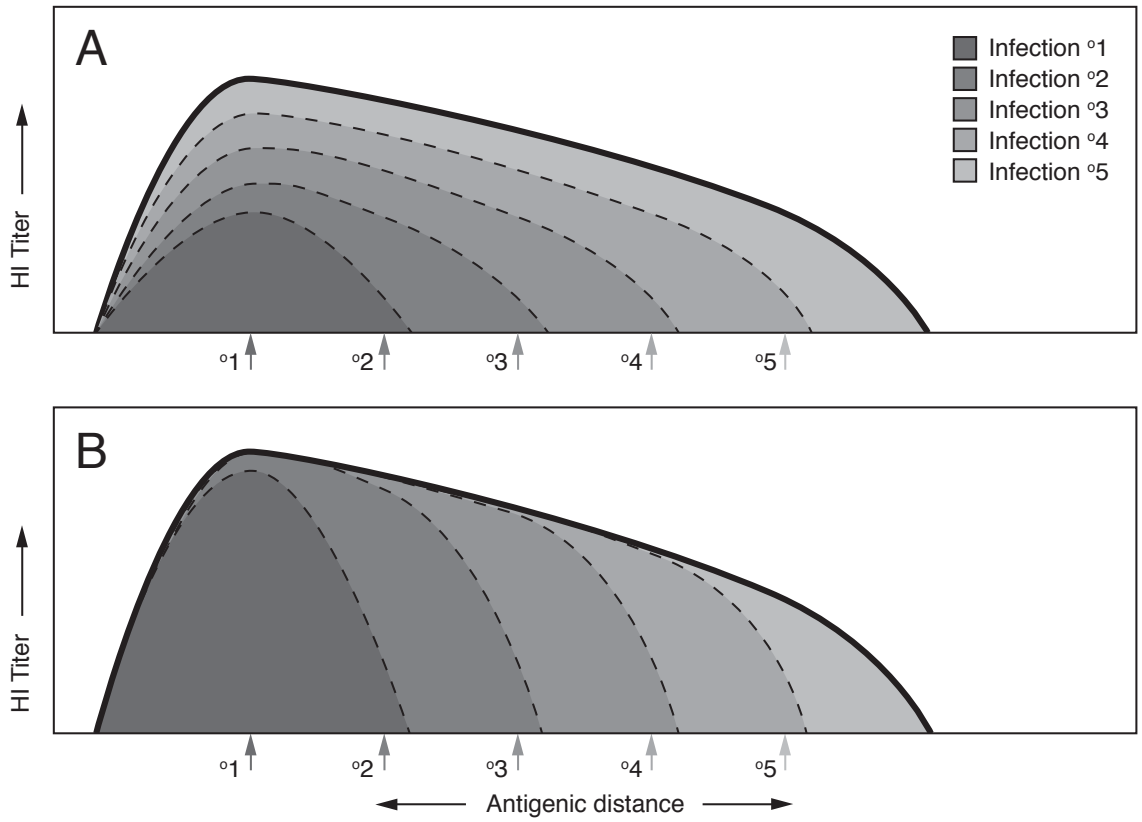



Figure S48: Schematic contrasting the building up of landscapes over time by a long-term and progressive reinforcement of all previous antibody titers (panel A), and by more localised responses, where the strength of the response is dependent on antigen trapping (panel B). Arrows represent the antigenic location of the strain involved in each subsequent infection. Shaded regions indicate the end result of each subsequent infection, as indicated by colors in the legend.

### 4.3 Conversion to a protection landscape

We envisage that antibody landscapes will be instrumental in the understanding and prediction of the evolution of antigenically variable pathogens. For example, the presence of antibodies with cross-reactivity against historical strains may explain why influenza does not regress, but appears to consistently evolve in antigenic directions that are consistent with ongoing avoidance of persistent immunological pressures (*2, 7*). More precise quantification of the extent of immune escape achieved through antigenic changes will be of paramount importance for studies aiming to predict the evolution of influenza, as population immunity is a critical part of the pathogen's fitness.

As an example, we converted the antibody landscape of an individual into a protection landscape based on a quantitative relationship between HI titers and clinical protection against influenza (*16*). The protection landscape in Figure S49 shows an individual's protection against the antigenically different influenza viruses. To convert the landscape, we use the relationship as determined in equation 2 by Coudeville *et al.* who described the functional form of relationship between HI titers and protection (the HI protection curve) as follows:

$$\text{Probability of protection} = 1 - \frac{1}{1+e^{\beta(\log(T_j)-\alpha)}}$$

where $T_j$ represents the HI titer and $\alpha$ and $\beta$ are parameters associated with location and steepness of the HI curve that they estimate to be 2.844 and 1.299 respectively.

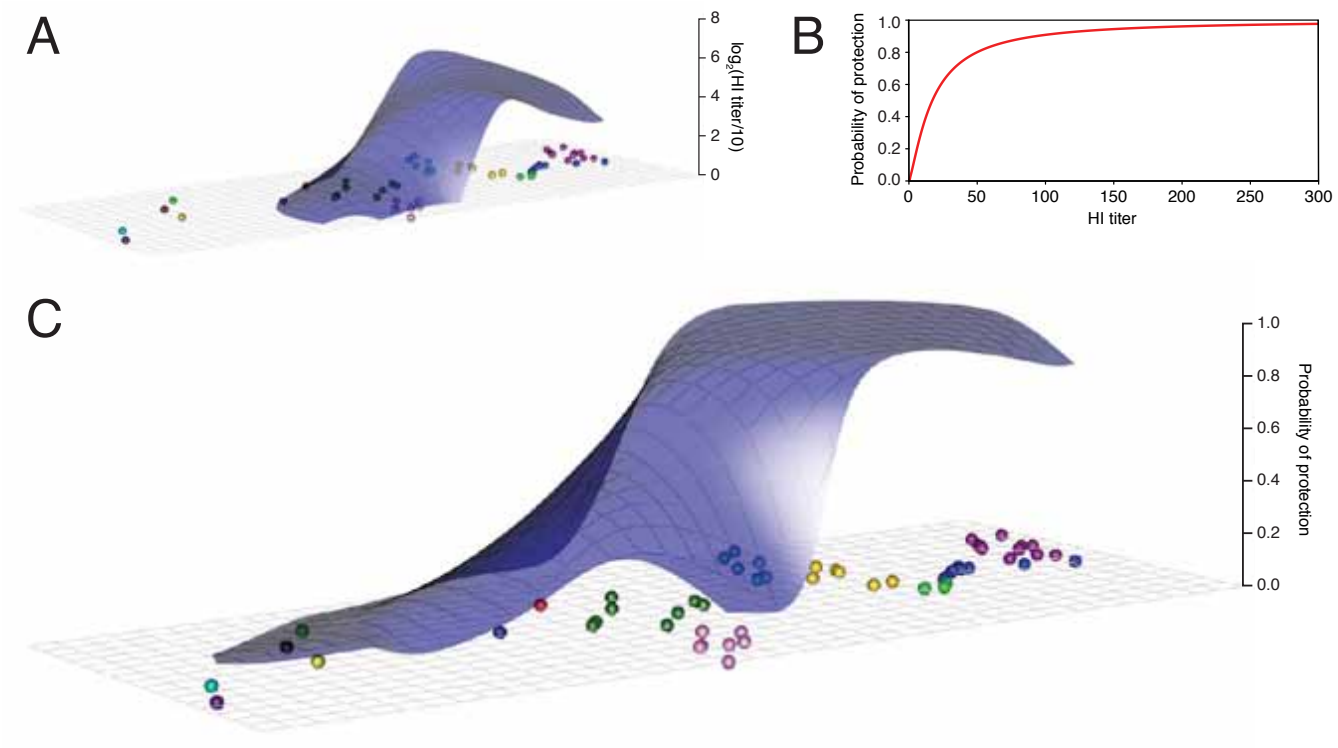



Figure S49: (A) An example antibody landscape can be converted via (B) the relationship between HI titer and clinical protection against influenza (*16*) to (C) a protection landscape, indicating the probability of protection.

If such a protection landscape can be made to represent the population's protection, based on HI immunity, against different influenza viruses, this would quantify a major determinant of viral evolution. Indeed, a population-based antibody landscape, via the protection landscape, provides a phenotypic landscape, as opposed to most other fitness landscapes, which are based solely on genotypic or theoretical considerations. Thus, the protection landscape is an important component of the pathogen's fitness landscape, where areas of lower population immunity would be expected to correlate to higher viral fitness and thus a greater likelihood of strain persistence (*34, 35*).