














Human Influenza A Virus Hemagglutinin Glycan Evolution Follows a Temporal Pattern to a Glycan Limit

 Meghan O. Altman,^a  Matthew Angel,^a  Ivan Košík,^a  Nídia S. Trovão,^{b,c}  Seth J. Zost,^d  James S. Gibbs,^a
 Lorenzo Casalino,^e  Rommie E. Amaro,^e  Scott E. Hensley,^d  Martha I. Nelson,^b  Jonathan W. Yewdell^a

^aCellular Biology Section, Laboratory of Viral Diseases NIAID, NIH, Bethesda, Maryland, USA

^bDivision of International Epidemiology and Population Studies, Fogarty International Center, NIH, Bethesda, Maryland, USA

^cGlobal Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, New York City, New York, USA

^dDepartment of Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

^eDepartment of Chemistry and Biochemistry and the National Biomedical Computation Resource, University of California, San Diego, La Jolla, California, USA

ABSTRACT Human antibody-based immunity to influenza A virus is limited by antigenic drift resulting from amino acid substitutions in the hemagglutinin (HA) head domain. Glycan addition can cause large antigenic changes but is limited by fitness costs to viral replication. Here, we report that glycans are added to H1 and H3 HAs at discrete 5-to-7-year intervals, until they reach a functional glycan limit, after which glycans are swapped at approximately 2-fold-longer intervals. Consistent with this pattern, 2009 pandemic H1N1 added a glycan at residue N162 over the 2015–2016 season, an addition that required two epistatic HA head mutations for complete glycosylation. These strains rapidly replaced H1N1 strains globally, by 2017 dominating H3N2 and influenza B virus strains for the season. The pattern of glycan modulation that we outline should aid efforts for tracing the epidemic potential of evolving human IAV strains.

IMPORTANCE Frequent mutation of its major antibody target, the glycoprotein hemagglutinin, ensures that the influenza virus is perennially both a rapidly emerging virus and a major threat to public health. One type of mutation escapes immunity by adding a glycan onto an area of hemagglutinin that many antibodies recognize. This study revealed that these glycan changes follow a simple temporal pattern. Every 5 to 7 years, hemagglutinin adds a new glycan, up to a limit. After this limit is reached, no net additions of glycans occur. Instead, glycans are swapped or lost at longer intervals. Eventually, a pandemic replaces the terminally glycosylated hemagglutinin with a minimally glycosylated one from the animal reservoir, restarting the cycle. This pattern suggests the following: (i) some hemagglutinins are evolved for this decades-long process, which is both defined by and limited by successive glycan addition; and (ii) hemagglutinin's antibody dominance and its capacity for mutations are highly adapted features that allow influenza to outpace our antibody-based immunity.

KEYWORDS influenza, viral evolution, hemagglutinin, immune evasion, glycosylation

Seasonal human influenza A virus (IAV) annually hospitalizes millions of people and kills hundreds of thousands (1). IAV remains endemic despite high levels of adaptive immunity in nearly all humans after early childhood. IAV persists due to selection of mutants with amino acid substitutions in the viral hemagglutinin (HA) glycoprotein that enable escape from neutralizing antibodies allowing periodic reinfection of adults with the same subtype. HA is a homotrimeric glycoprotein with a globular head domain resting atop a stem that anchors HA to the virion surface. The head has a receptor binding site that attaches virions to cell surface terminal sialic acid residues, initiating

Citation Altman MO, Angel M, Košík I, Trovão NS, Zost SJ, Gibbs JS, Casalino L, Amaro RE, Hensley SE, Nelson MI, Yewdell JW. 2019. Human influenza A virus hemagglutinin glycan evolution follows a temporal pattern to a glycan limit. *mBio* 10:e00204-19. <https://doi.org/10.1128/mBio.00204-19>.

Editor Diane E. Griffin, Johns Hopkins Bloomberg School of Public Health
This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply.
Address correspondence to Meghan O. Altman, meghan@altmans.org, or Jonathan W. Yewdell, jyewdell@niaid.nih.gov.

This article is a direct contribution from a Fellow of the American Academy of Microbiology. Solicited external reviewers: Ian Wilson, The Scripps Research Institute; Cornelis de Haan, Utrecht University.

Received 25 January 2019

Accepted 15 February 2019

Published 2 April 2019

infection. Ab binding to the head neutralizes virus by blocking attachment. Understanding Ab-driven HA evolution is essential to improving influenza vaccination, which currently offers only moderate protection from infection (2, 3).

As newly synthesized HA enters the endoplasmic reticulum of IAV-infected cells, oligosaccharides are often, but not always, attached to asparagine (Asn; N) residues present in the motif Asn-X-serine/threonine-Y, where X and Y represent any other amino acid except proline. Attached glycans can be high-mannose glycans or larger, more complex branched glycans (4, 5). HAs possess multiple highly conserved glycans on the stem domain crucial for HA folding and oligomerization (6). In birds, HA heads typically remain minimally glycosylated; it has been carefully noted that evolving HAs in humans, by contrast, accumulate over time (7–9) and occasionally swap head glycosylation sites (8, 10, 11).

Head glycans are known to promote viral fitness by shielding virus from antibody binding and by tuning receptor affinity and specificity (12–15). The importance of glycan-mediated shielding was highlighted recently by a study that revealed the inability of ferrets and humans exposed to an otherwise well-matched, current H3N2 vaccine strain lacking the most recently added glycan to produce antibodies that robustly neutralized glycosylated virus (16). Despite their clear role in antibody evasion, glycan changes are generally not correlated with changes between antigenic clusters (17).

This is the case in part because glycan addition typically lowers HA receptor avidity (18), complicating binary assays such as those evaluating hemagglutination inhibition (HI) (19). Further, antigenic clusters are defined by HI assays performed using ferret antisera, but HA head glycans are selected to escape neutralization by human antibodies specific for antigenic sites that, typically, are weakly targeted by ferret antibodies and that may not inhibit viral attachment (20). Consequently, despite glycan evolution being a major contributor to antigenic escape, it is often left unaccounted for in studies using HI-based analytic tools such as antigenic cartography (21, 22).

Despite selective advantages offered by antigenic shielding, glycan addition can also exact a high fitness cost for viruses by deleteriously altering receptor binding, enhancing viral neutralization by innate immune lectins, or increasing stress in the endoplasmic reticulum during translation (19, 23–27).

RESULTS

Empirical analysis of relative HA sizes over time. Studies of HA glycan evolution rely nearly exclusively on bioinformatic predictions of glycan addition. NetNGlyc (28) predicts N-linked glycosylation (5), and servers such as GlyProt can be used to filter out sites by considering structural constraints such as surface accessibility, but neither provides either a guarantee that glycans were added or information regarding added glycan structure. We therefore used gel electrophoresis to monitor glycan addition to HAs from 72 egg-grown H1N1 and H3N2 strains as determined by decreased HA migration (Fig. 1; see also Fig. S1 to S3 and Table S1 in the supplemental material and File S1 [http://downloads.misms.net/Publications/Glycosylation_mBio/index.html]).

Growing all viruses in eggs ensured a standardized viral output. Our choice of assay is associated with several potential experimental artifacts; namely, glycosylation patterns can differ between species, and egg adaptation of human viruses can lead to mutations in HA. Further, amino acid differences between HAs could also affect electrophoretic shift independently of glycosylation. PNGase treatment of viruses indicated that this effect was minimal compared to glycosylation-mediated size changes (Fig. S2D and E). Despite the potentially confounding influence of our experimental choice, with the large sample size, we found that HA migration strongly correlated with the number of computationally predicted head glycosylation sites (Fig. 1A). H1 adds an apparent 3.3 ± 0.2 kDa ($R^2 = 0.93$) per glycan, suggesting that the majority of such glycans are highly branched. The average H3 glycan, by contrast, is smaller (2.5 ± 0.2 kDa, $R^2 = 0.44$), consistent with greater levels of addition of simpler glycans. The size difference between H1 and H3 HA glycans is consistent with mass

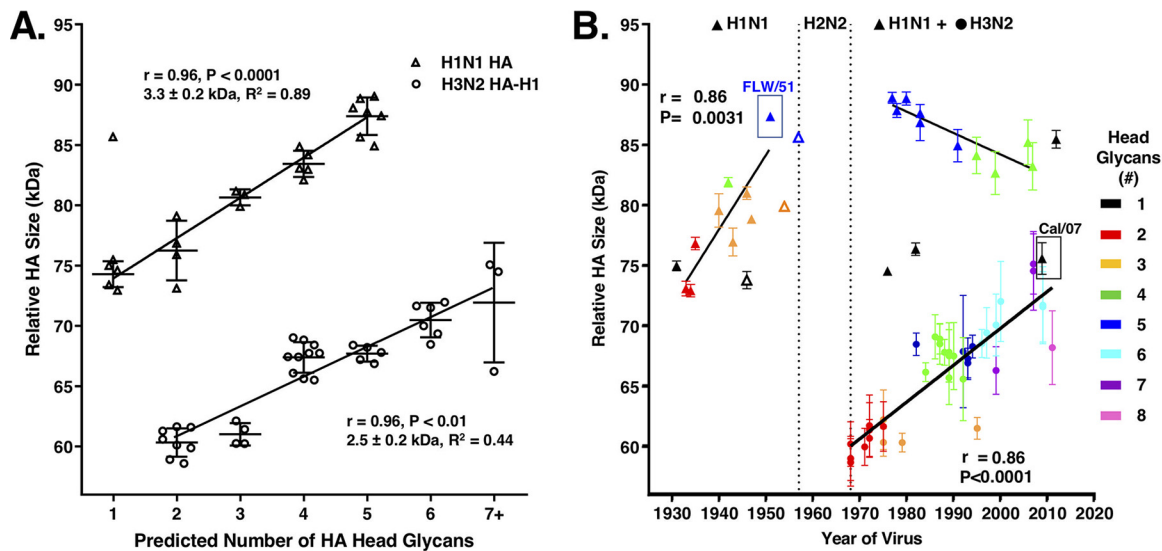


FIG 1 Quantitative survey of HA glycan numbers in IAV. (A) Relative HA sizes, measured by migration rate through a gel, correlate with predicted numbers of head glycans. H1N1s (triangles) add 3.3 ± 0.2 kDa per glycan (or 13.1 kDa in total) while increasing from 1 to 5 glycans. H3N2s (HA1; circles) add 2.5 ± 0.2 kDa per glycan (or 12.5 kDa in total) while increasing from 2 to 7 glycans. Each point represents the mean HA size for one virus. Error bars represent standard deviations of the means. (B) HA sizes over time for H1N1s (triangles) and H3N2s (HA1; circles) reveal regular glycan addition. Glycan outliers and strains not included in analysis (open triangles) are addressed in Materials and Methods. Strains are color-coded to match glycan numbers (Fig. S3). A glycan is added on average every 5.4 years for sH1N1 and every 8.1 years for H3N2. The strain most closely related to the sH1N1 virus reintroduced in 1977, A/Fort Leonard Wood/1951, is boxed and labeled in blue. The pH1N1 strain, A/California/07/2009, is boxed and labeled in black. Data representing the Pearson coefficient of correlation (r) and its P value are shown on both graphs. Error bars represent standard errors of the means of data from $n = 3$ to 5 blots.

spectrometry characterizations showing that H1 glycans skew to larger, more complex structures whereas the majority of H3 glycans at several sites are high-mannose glycans (5, 12).

Importantly, the total numbers of glycans and glycan mass added to the head during antigenic drift in humans are similar between subtypes: H1 added 4 glycans, or 13.1 kDa, for a maximum of ~ 16.5 kDa of glycans per monomer, and H3 added 5 glycans, or 12.5 kDa, for a maximum of ~ 17.5 kDa of glycans per monomer. These values are consistent with the idea of an upper limit of glycan shielding for H1 and H3 HA head domains, although it remains to be seen which specific features of glycosylation, such as number, mass, or type, are important for determining this limit.

H1 HA glycan evolution. Plotting relative HA size versus time reveals the history of seasonal H1N1 (sH1N1) glycosylation (Fig. 1B, triangles). First-wave sH1N1 (1933–1957) HAs behaved differently from the second-wave sH1N1 representatives (1977–2008). During the first wave, relative HA size increased by an average of 3.3 kDa, or one glycan, every 5.4 years (slope = 0.61 kDa/year, $R^2 = 0.86$).

As expected, the sH1N1 strains that reemerged after the 1977 reintroduction of sH1N1 exhibited relative HA sizes similar to those of the most related strain, A/Fort Leonard Wood/1951 (inside blue box, Fig. 1B). Rather than increasing in number as had occurred previously, electrophoresis confirmed the NetNGlyc prediction that these 5 HA-head glycan strains were maintained until they decreased to 4 glycans after 1991.

Indeed, HA sequences predicted to have 6 head glycans were highly unusual over the century of H1 evolution (1918 to 2017), except in 1986 and 1987, when they accounted for $>50\%$ of H1 strains (Fig. S4A, black). Rather than adding glycans, sH1N1 strains that reemerged after 1977 exhibited glycan adjustments spaced at longer intervals of 10.3 ± 0.8 years (Fig. 2). In 1986, N54 pandemic H1N1 (pH1N1) (the numbering convention is shown in Table S1) replaced N155, as N127 swapped with N125. Around 1997, N269 was lost. Strains with the 4 remaining glycans circulated for 11 years before becoming extinct during the 2009 pH1N1 pandemic. During at least 33/35 seasons after the glycan limit was reached, sH1N1 failed to evolve a widely circulating strain with 6 glycans. This history is consistent with a functional glycan limit.

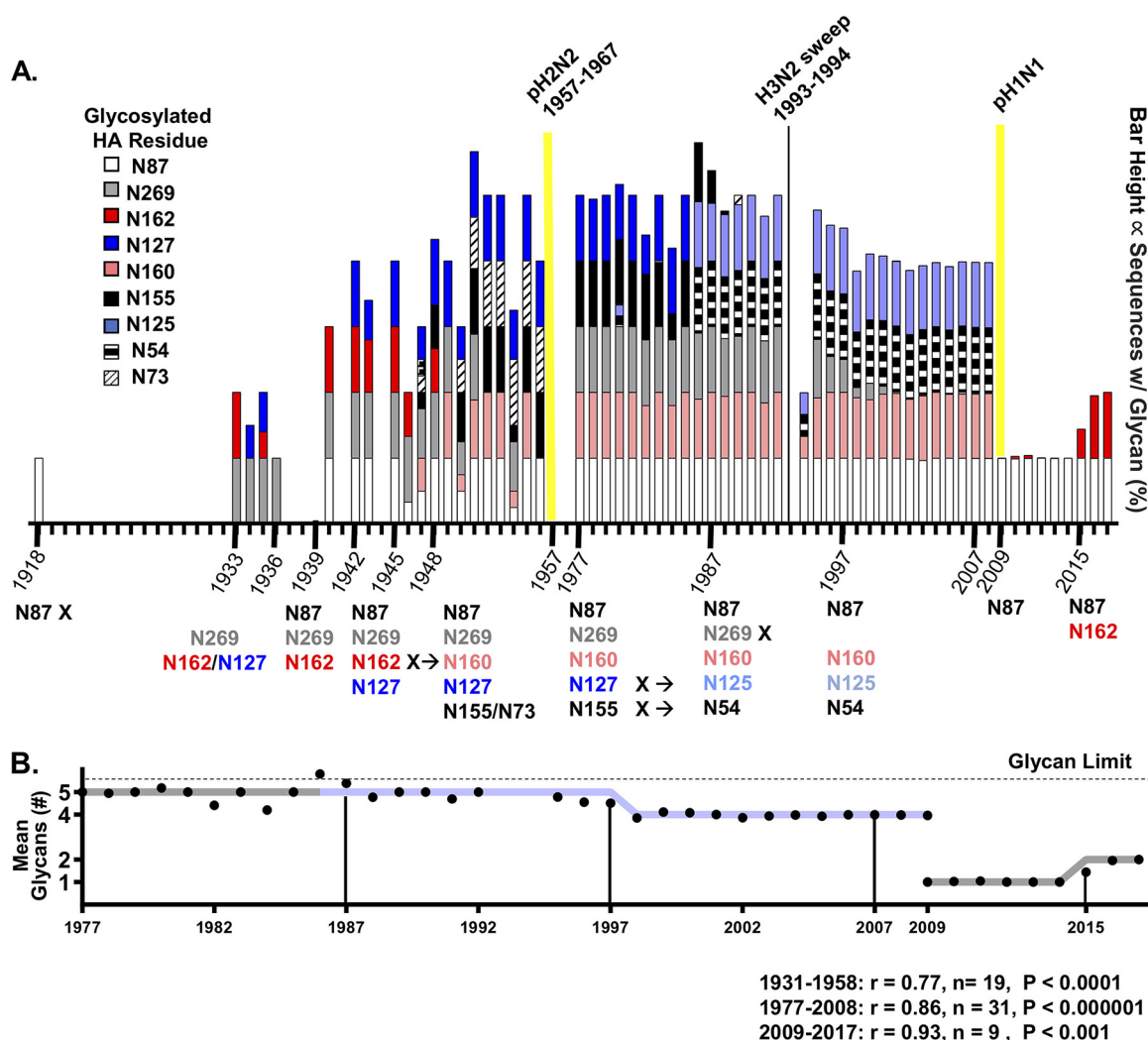


FIG 2 H1 glycan evolution, 1918 to 2017. (A) Stacked bars correspond to the percentages of all human H1N1 sequences from a given year containing a glycan listed in the legend. Dates of glycan transition are indicated on the timeline at the bottom, along with the glycosylated residues. Years with no data were left blank. Sequence counts for each year are shown in Fig. S4A. Pandemics are denoted by yellow lines. (B) Mean predicted numbers of glycans among all sequences from each year are shown as black circles, which are displayed atop the idealized pattern (gray line; the light blue line indicates glycan swap from sequences containing N127/N155 to sequences containing N125/N54). Black vertical lines denote 11-year intervals after the glycan limit was reached or 6-year intervals during glycan addition. Data representing Pearson's coefficient of correlation (r) with sample number (n) and P value, from two-tailed Student's t test comparing the pattern versus the sequence data are shown at the bottom.

H2 HA glycan evolution. Despite being experimentally possible, during their 11-year era (1957 to 1967), human H2N2 never added or swapped any glycans.

The H2N2 HA head has two highly conserved overlapping glycans with the sequence NNTS, where either, but not both of the Asns can be glycosylated (29). While it remains unknown which Asn (or which mixture) is glycosylated in H2 HA, the kinetics of NXT glycosylation are faster than the kinetics of NXS glycosylation (30) and overlapping sites in yeast, mice, pH1N1, and H5N1 prefer NXT to NXS (31–34).

Selection of H2N2 strains with additional head glycans outside the overlapping site occurs readily during *in vitro* antibody escape. These multiglycan H2 HAs, however, exhibit drastically lower levels of cell fusion and receptor binding (29).

H3 HA glycan evolution. As with H1 HAs, H3 HAs exhibit regular intervals of glycan evolution, with one exception (Fig. 3). The original (1968) pandemic H3 strains had 2 head glycans. Four times since 1968, glycans were added to H3 HA at regular intervals

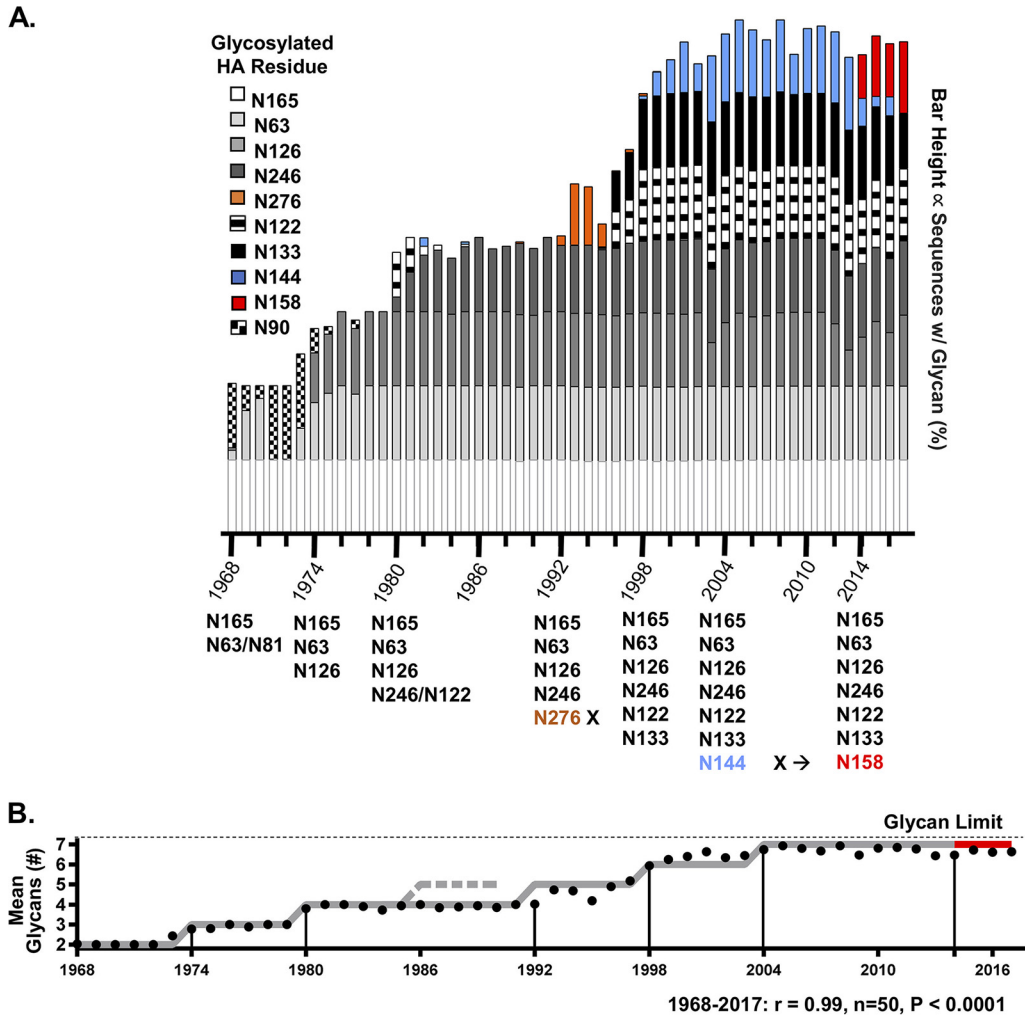


FIG 3 H3 glycan evolution, 1968 to 2017. (A) Stacked bars correspond to the percentages of all human H3N2 sequences from a given year containing a glycan listed in the legend. Dates of glycan transition are indicated on the timeline at the bottom along with the glycosylated residues. Sequence counts for each year are shown in Fig. S4B. Mean predicted numbers of glycans among all sequences from each year are shown as black circles, which are displayed atop the idealized pattern (gray line; the red line indicates glycan transition from sequences containing N144 to sequences containing N158). The dashed gray line indicates when N276 would have been added if it had followed the pattern. Black vertical lines denote 11-year intervals after glycan limit was reached or 6-year intervals during glycan addition. Data representing Pearson's coefficient of correlation (r) with sample number (n) and P value from two-tailed Student's t test comparing the pattern versus the sequence data are shown at the bottom.

of 5 to 7 years (N126 by 1974, N246/N122 by 1980, N133-plus-N122 replacement of N276 by 1998, and N144 by 2004) (Fig. 4). This timing is similar to that corresponding to the average rate of glycan addition (one glycan every 5.4 years) seen by SDS-PAGE for sH1N1 from 1933 to 1951.

With the addition of the 7th head glycan in 2004, H3 HA had increased in size by 12.5 kDa from the original 1968 strain. We hypothesize that H3N2 reached the functional glycan limit at this point (Fig. 3B). It is apparent from sequence data that H3 HAs with >7 head glycans lose the competition for human circulation. While 8-glycan strains circulate, they represent $<1\%$ of H3N2 sequences in any year (Fig. S4B). Similarly to the second-wave H1 viruses, 12 to 15 years after reaching the glycan limit, a glycan swap occurred, as the glycan at residue N144 was lost and a new residue at N158 was gained (Fig. 3B; see also Fig. S5B).

There was one exception in the 5-to-7-year-interval pattern for H3 glycan addition. Addition of a fifth glycan, at residue 276, occurred 12 to 13 years after the previous glycan addition (Fig. 3A, orange). In 1993, robust strains with N276 swept the globe,

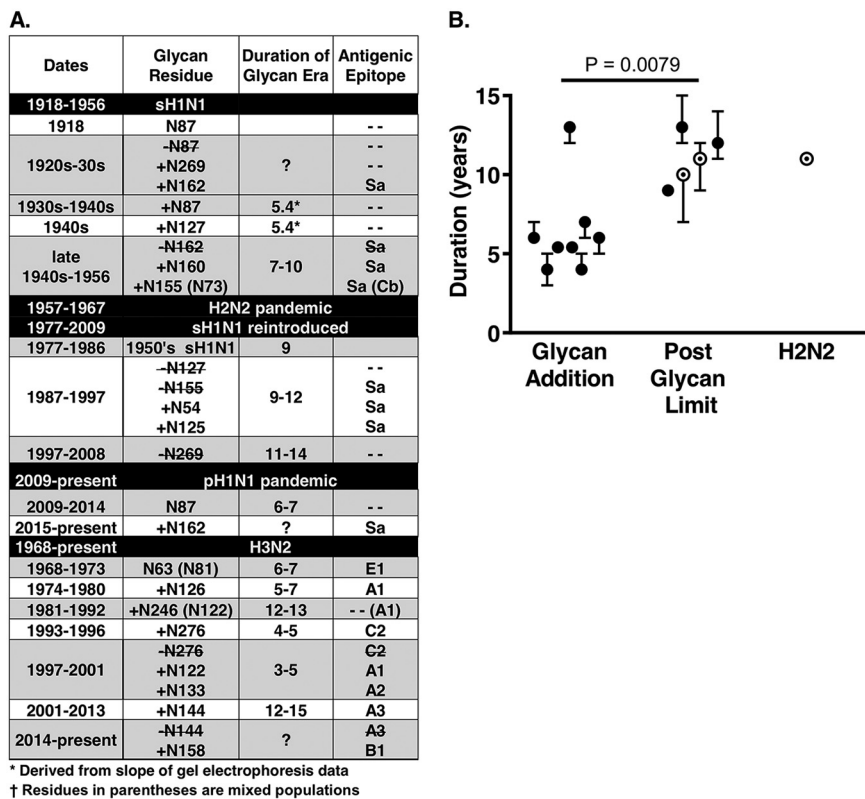


FIG 4 Quantitative analysis of glycan eras. (A) The duration (in years) of each glycan era corresponds to the interval when >10%, >50%, or >80% of sequences had the new glycan arrangement. The data shown for the duration of the glycan era correspond to the range of these three dates. (B) The durations of glycan eras ended by glycan addition were, on average, half as long as those seen after the glycan limit was reached (Student's two-tailed *t* test, $P = 0.0079$). Eras ended by pandemic replacement are denoted by open circles.

dominating sH1N1 strains, before disappearing. N276 is unique as the glycan-bearing Asn residue is immediately adjacent to a structurally important disulfide-bonded cysteine (C275) which defines the HA head-stem border.

pH1 HA follows the sH1 glycosylation pattern. The initial 2009 pH1N1 strains had a single predicted head glycan (N87). As expected, the HA of a representative early strain, A/California/07/2009 (Fig. 1B, inside black box), comigrated with strains from the 1930s and with two other swine-origin human strains, A/New Jersey/8/1976 and A/Memphis/4/1982, with a single predicted head glycan (Fig. 1B, red triangles). From SDS-PAGE, we calculated that seasonal H1N1 added a glycan every 5.4 years before reaching the glycan limit. If pH1N1 were to follow the same trajectory, we predicted that it would add a glycan over the 2015 to 2016 season. Analysis of sequences from this period revealed that, indeed, a second glycan site arose through an S-to-N change at residue 162 over the 2015–2016 season (Fig. 5A and B).

Two changes are necessary for complete glycosylation of N162. N162 viruses were present at low frequencies in each season since pH1N1 introduction (Fig. 5A and B). Viruses with this mutation did not selectively sweep the human pH1N1 strains until the 2015–2016 season (Fig. 5; see also Fig. S5A). All of these strains, unlike previous N162 strains, also had an I216T mutation.

Residue 162 is in an antigenic site ("Sa") (35) that is immunodominant in many contemporary human Ab responses (36). Prior to the N162 glycan sweep during the 2013–2014 season, pH1N1 was swept aside by strains that were antigenically distinct due to a K163Q mutation co-occurring with an A256T mutation (36, 37). Surprisingly, through reverse-genetics manipulation of the A/California/07/2009 strain, we found that, despite their having the amino acid sequence necessary for glycosylation, neither

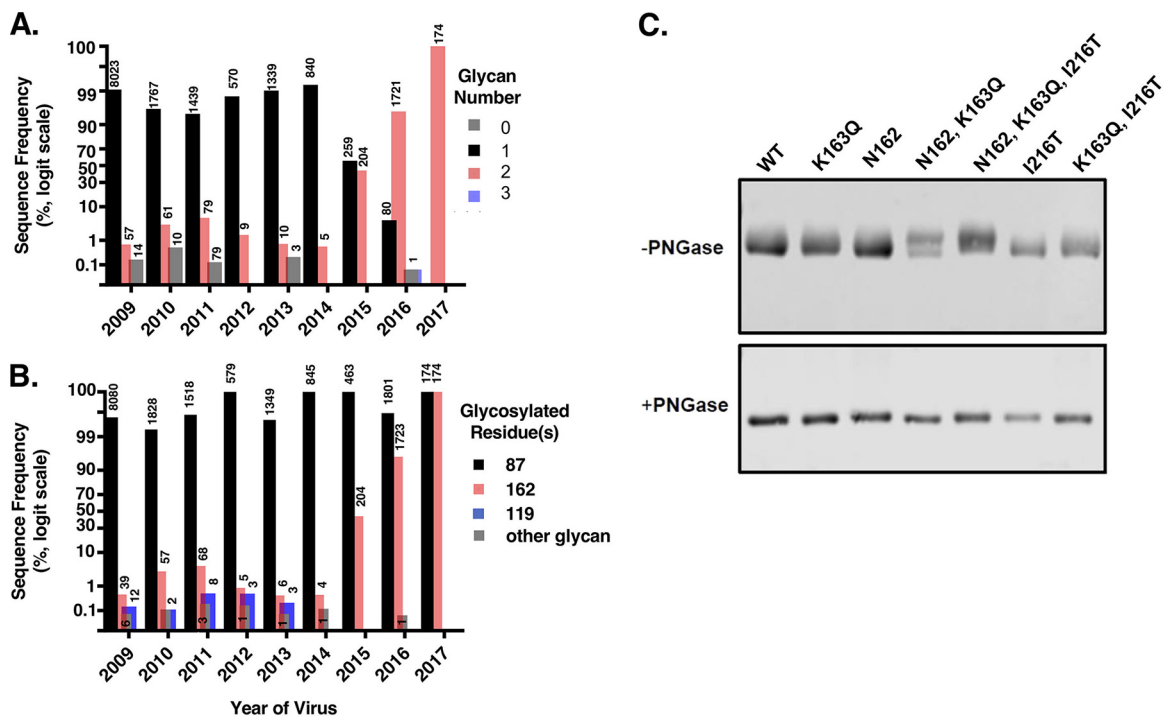


FIG 5 pH1N1 glycan addition predicted from gel migration data. (A) Frequency of human pH1N1 sequences in FluDB determined by the number of predicted glycans in each calendar year. The number of sequences represented by each bar is shown above the bar. Until January 2015, >90% of sequences had 1 glycan (black). Starting in 2015, strains with 2 glycans swept the globe (red). (B) Frequency of individual residues glycosylated in human pH1N1 sequences. N87 predominates throughout (>95% pH1N1). N162 (red) and N136 (blue) appeared at low frequencies until 2015, when strains having both N162 and N87 dominated. N162 strains were found in every year. Rarely, in addition to N87, a residue other than N162 or N136 was found to have a potential glycosylation site (gray). Unlike N162, a single mutation with N136 alone is sufficient for glycosylation (14). (C) HA Western blots of A/California/07/2009 mutants with permutations of N162, K163Q, and I216T with and without PNGase treatment to remove glycans. The single mutant with N162 alone did not shift in size, indicating that it was not detectably glycosylated. While a portion of the double mutant N162/K163Q shifted, a complete shift, indicating majority glycosylation, was seen only with the triple mutant N162/K163Q/I216T.

the N162 single mutants nor the N162/K163Q double mutants added a glycan to HA in virions. Rather, three mutations, namely, N162 itself, K163Q, and, crucially, I216T, were required for complete glycosylation of the 162 site (Fig. 5C).

Interestingly, in H3 strains that also possess a N162 glycan, the residue equivalent to residue 216 in pH1N1 is also a small polar amino acid, serine, suggesting strong epistasis between S and T at position 216 and glycosylation at position 162. This is puzzling, since as a cotranslational process, N-linked glycosylation should not be controlled by a residue 54 positions downstream from the glycosylation site (38). Could it be that residue 216 influences the fitness of residue 162 glycosylation at the level of trimer structure/function? To calculate approximate distances between the residue 162 glycan and the 216 residue, we created a homology model of A/Michigan/45/2015 HA with a likely triantennary glycan at N162 (Fig. 6). The model shows that the side chain of 216 is highly exposed and faces the glycan, consistent with an epistatic interaction. It must be added, however, that in a static model, the distance of ~ 9 Å between the position 162 glycan and residue 216 is inconsistent with steric or polar interactions with T216 (< 3.5 Å). Possible explanations include the following: (i) the attached oligosaccharide is larger than the glycan modeled; (ii) the conformational breathing of the HA brings the glycan into proximity to residue 216; and (iii) residue 216 epistatically interacts with the glycan (or with glycosylation itself) on the monomer rather than on the trimer.

Phylogeographic reconstruction of glycan sweep. A phylogenetic analysis of all HA sequences from pH1N1 viruses collected globally during 2009 to 2017 revealed that N162/I216T viruses emerged repeatedly during 2009 to 2014, particularly in North

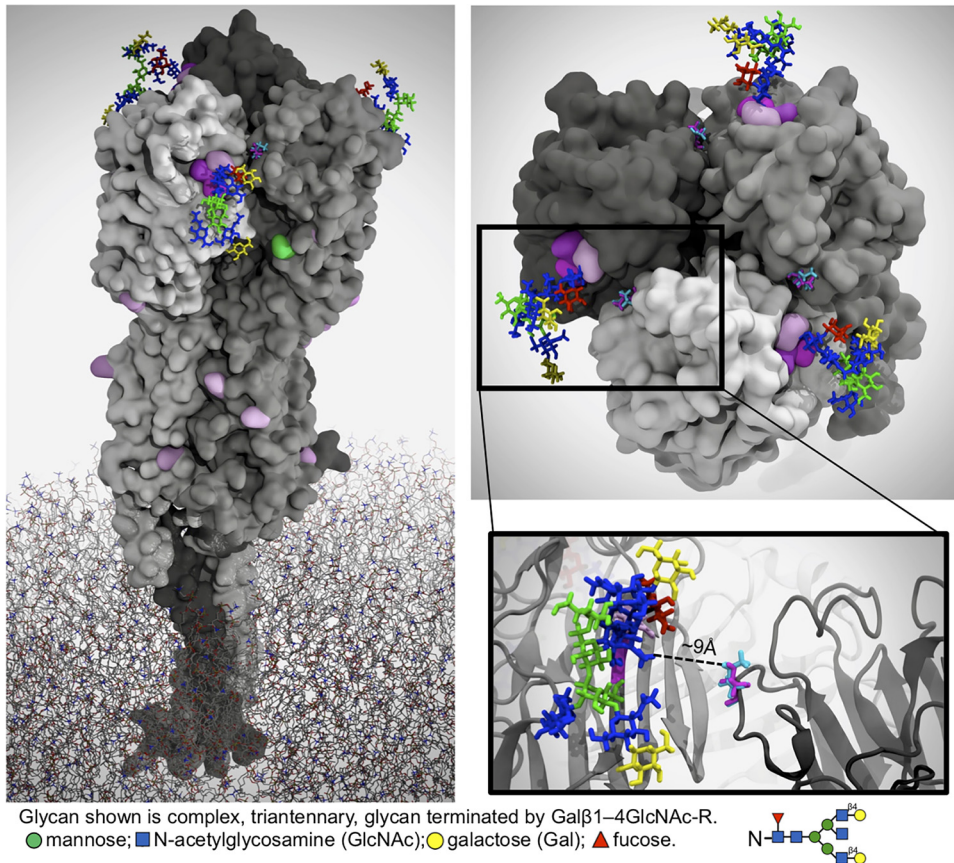


FIG 6 HA model highlighting key HA residues. Complete A/Michigan/45/2015 HA with an attached N162 glycan, embedded in a lipid bilayer, is shown from the side (left) and top (right). Three residues defining the N162 glycosite have undergone recent mutations, and their surfaces are shown in pink as follows: S162N, 2016–2017; K163Q, 2013–2014; and S164T, 2017–2018. The original I216 mutation (cyan) the and compensatory T216 mutation (magenta) stick side chains are overlaid across the monomer from the glycan. The surface of a slightly earlier antigenic site mutation, S84N, is indicated in green. All additional glycosylated Asn residues on HA are shown in pink. The inset shows a magnified view of the mutated residues with ~ 9 Å of distance between the Thr hydroxyl group and the nearest GlcNAc hydroxyl group.

America and Europe (Fig. 7A, blue spikes), but did not predominate and were quickly replaced by other strains (Fig. 7A; see also File S12 and S13 at http://downloads.misms.net/Publications/Glycosylation_mBio/index.html). Previous studies have identified 2014 as a turning point in pH1N1 evolution, when the pH1 phylogeny became more ladder-like, with a single dominant lineage, presumably owing to a high proportion of the global population being exposed to pH1N1 by that time (11).

Our findings are consistent with this observed timing of immune-driven selection, as the N162/I216T strains rapidly swept to dominance globally beginning in late 2014 and represented 100% of pH1N1 viruses in most regions by 2017. The sharp rise of N162/I216T strains to dominance during 2015 occurred simultaneously in multiple regions (Fig. 7A; see also Fig. S6A), making it difficult to infer the most likely spatial origin of these strains and resulting in low posterior probabilities for all regions (Fig. 7B). In contrast, the geographical origins of the S84N substitution is strongly supported in South Asia, where S84N strains gradually increased in frequency during 2013 to 2015 before becoming dominant in other regions (Fig. 7; see also Fig. S6B and Fig. S7A), providing a distinct spatial pattern for the phylogeny and high posterior probabilities (0.74 in South Asia). Similarly to N162/I216T, the S84N substitution also was present in pH1N1 viruses during earlier years of the pandemic but also died out quickly (Fig. 7A, gray spikes).

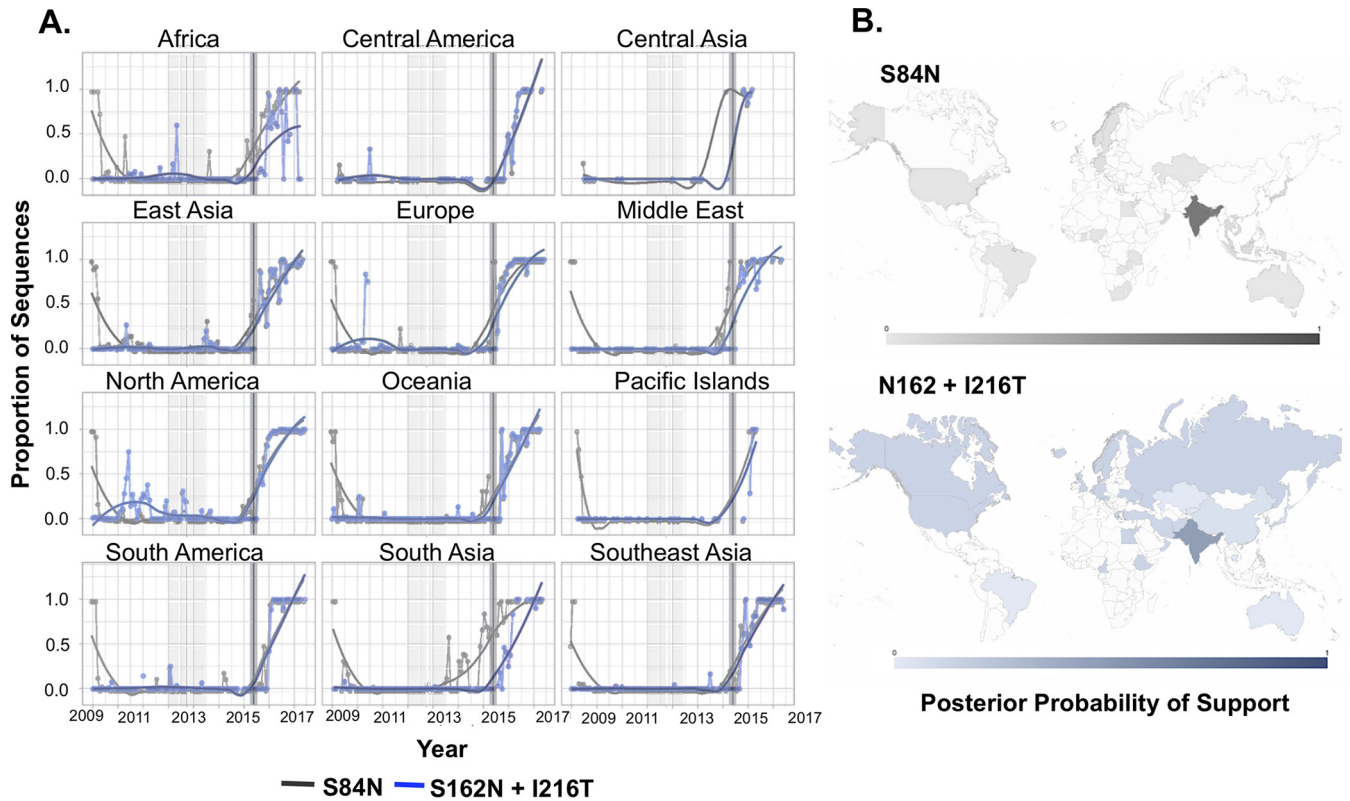


FIG 7 Global emergence of N162/I216T strains. (A) Proportions of all globally available H1 sequences from pH1N1 viruses collected in 2009 to 2017 from the GISAID database with the N162/I216T substitutions are indicated for 12 geographical regions over time (in Africa, Egypt, Cameroon, and Ethiopia; in Central America, Costa Rica; in Central Asia, Kazakhstan and Kyrgyzstan; in East Asia, Japan, China, Hong Kong, Taiwan, and South Korea; in Europe, Iceland, Austria, Netherlands, Germany, Russia, Denmark, Norway, United Kingdom, Ukraine, Greece, Sweden, and Belgium; in the Middle East, Iran, Oman, Bahrain, Jordan, Turkey, Afghanistan, and Israel; in North America, the United States and Canada; in Oceania, Australia; in the Pacific Islands, Hawaii; in South America, Brazil; in South Asia, India, Nepal, Bangladesh, and Pakistan; and in Southeast Asia, Singapore, Philippines, and Cambodia). The 95% HPD intervals for mean time to the most recent common ancestor (tMRCA) are indicated by vertical blue or gray shadows. (B) The posterior probability support for the geographical origin of the major clade of N162/I216T and S84N viruses is indicated by the shade of the country (light blue = low posterior probability, dark blue = high posterior probability), inferred from the MCC tree. This figure is reproduced with separated proportions of sequences panels in File S11 (http://downloads.mims.net/Publications/Glycosylation_mBio/index.html).

DISCUSSION

Through empirical analysis of HA glycosylation, we revealed regular patterns of glycan evolution that have followed decades-long trajectories. Initially, a strain from the animal IAV reservoir with minimally glycosylated HA establishes worldwide circulation in humans. At regular 5-to-7-year intervals, head glycans are added to HA. After a functional glycan limit is reached, head glycans are swapped at longer intervals until pandemic replacement occurs (Fig. 4). Glycan changes have occurred with discrete, globally rapid, selective sweeps of IAV, as well as with slower increases over multiple seasons (see Fig. S5 in the supplemental material).

Several issues surrounding this pattern will require further study. For instance, if the major role of glycosylation is to shield antigenic sites, loss or swapping of a glycan would presumably open a previously covered antigenic region against which older individuals might be protected. Determining the impacts of glycan addition versus glycan loss or swap will be important. Considering both H1 and H3 HA glycan evolution, glycan loss occurred in all 3 glycan changes after their glycan limits were reached (for H1N1, loss of N127 and N155 around 1987 and loss of N269 around 1997; for H3N2, loss of N144 around 2014).

The pattern described in this paper derives from 13 glycosylation changes and 3 subtype extinctions over the past 100 years. As these events are rare, the data set is limited and even within the limited data, the pattern contains inconsistencies. The biggest of these inconsistencies is the addition of N276 to H3N2 in 1992 to 1993 after

12 to 13 years rather than 5 to 7 years. In the HIV env glycoprotein, glycans are enriched at the C-terminal residues of disulfide bonds, but absent from residues on the N-terminal side. We speculate the N-terminal glycan at N276, adjacent to C277, may have reduced the fitness of these strains and their competitiveness against future strains lacking N276 (57).

These and other concerns leave open the issue of whether IAV will maintain the patterns outlined here in the near future. For instance, will pH1N1 rapidly add a third glycan in the next few years? Will H3N2 once more swap (or lose) glycans or will it go extinct in the next decade? Or, instead, will it surpass our proposed glycan limit by adding an eighth glycan?

What is clear is that glycosylation sites in both pH1N1 and H3N2 are under intense selective pressure. For instance, pH1N1 HA mutation S164T swept the globe over 2017 to 2018, altering the N162 glycosite motif from NXT to NXS (Fig. S5A). The kinetics of NXT glycosylation are known to be faster than those of NXS glycosylation, suggesting that this mutation could facilitate glycosylation at N162. This also means that between 2013 and 2018, all three residues composing the N162 glycosylation site had mutated (SKS to NQT). For H3N2, two new clades have recently begun to circulate with either a T135K or T135N mutation. Both mutations remove the glycan at N133; however, the T135N mutation also creates a new glycosylation site at N135 (Fig. S5B). With either 6 or 7 glycans, these strains remain under the glycan limit.

Experimentally, it has been shown that H1 HAs can accommodate at least 8 head glycans while maintaining *in vitro* fitness (39). However, sH1 circulated for over 3 decades after reaching 5 glycans without fixation of a sixth glycan. Interestingly, at 9 to 10 years after reintroduction in 1977, sH1N1 strains with 6 glycans did briefly circulate before being replaced by strains with only 4 glycans. These findings point to *in vivo* glycan-based fitness costs involved in maintaining the glycan limit, possibly due to structural constraints on the HA protein, innate immune mechanisms, or alterations in virus binding to receptors found in human airways (40). However, if the pattern does indeed hold up over time, it requires explanation. We propose a simple albeit provocative rationale to account for this observation.

Every IAV season, swarms of closely related minor variants sample broad swaths of mutational space (41), including new glycan sites. In large, age-stratified populations, this constant sampling allows circulating strains to seasonally detect and quickly respond to changes in the immune context. In an antigenically naive population, an HA head with n glycans is fitter than an HA head with $n + 1$ glycans. As Ab pressure mounts, the fitness costs associated with adding a glycan are outweighed by the advantage conferred by antigenic shielding. We propose that a period of 5 to 7 years is required to provide sufficient herd immunity to enable the dominance of glycan addition mutants.

We also note that the average period of time during which a glycan addition has become dominant in human H1 or H3 strains is strikingly similar to the 5.3-year average for the emergence of norovirus genotype GII.4 variants over the past 30 years (42). While glycan evolution is not involved in creating antigenically distinct GII.4 norovirus variants, the similar kinetics of these two globally drifting viruses may reflect broad similarities in the rates at which global humoral immunity builds to enable selection of escape mutants with higher fitness costs.

On the basis of our work, using biochemistry, immunology, and phylogenetics, we can attempt to reconstruct the 2016 selection of the N162 glycan addition. Antibody pressure on the Sa antigenic site first selected for the K163Q escape mutation, which replaced other pH1N1 strains. Several years of immune pressure against this new Sa site led to co-selection of the N162 glycan and I216T, which increases N162 glycosylation.

In humans, the HA head of both the H1 and H3 subtypes contains 5 or so potentially immunodominant, mutationally plastic, largely antigenically independent epitopes. Evolution among these epitopes over decades facilitates repeated infection of adults with the same IAV subtype. The HA head's antigenic dominance (43), in combination with the glycosylation patterns outlined in this paper, has conferred a fitness advantage

for long-term IAV persistence in large, age-stratified populations whose members are reliant on anti-HA head antibody-mediated IAV immunity, i.e., humans. The patterns in glycan evolution that we have observed likely increase viral fitness by subverting antibody-based immunity over decades-long time scales in these populations. Taken together, the data indicate that H1 and H3 HAs may be adapted for optimized antigenic drift or, put a different way, that HA has evolved to drift. If true, similarities among sH1, pH1, and H3 glycan changes highlight the possibility that rough blueprints for the pattern that we observe are potentially stored in the fitness landscape of the HA protein itself.

MATERIALS AND METHODS

Virus preparation and purification. We propagated each virus by infecting five 10-day-old embryonated chicken eggs with 100 μ l of viral stock diluted 1:2,000 in phosphate-buffered saline (PBS). Infected eggs were incubated at 34.5°C and 65% humidity for ~41 h. We first clarified the allantoic fluid (AF) by centrifugation at 4,500 $\times g$ for 20 min and then pelleted the virus by centrifugation for 2 h at 26,000 $\times g$. After incubating the viral pellet in 1.7 ml PBS overnight at 4°C, we layered the virus onto a discontinuous 15%-to-60% sucrose gradient and spun it at 34,000 $\times g$ for 2 h. We collected virus at the gradient interface and spun it a final time in PBS at 34,000 $\times g$ for 2 h. Virus was resuspended in 100 μ l PBS and stored as milky opalescent suspensions in long-term storage at 4°C. Hemagglutination titers for each virus were measured from neat AF. A complete list of viruses and their origin is available in File S1 at http://downloads.misms.net/Publications/Glycosylation_mBio/index.html. The total protein content of each virus preparation was determined by a Coomassie blue-based assay performed per the instructions of the manufacturer (Bio-Rad; DC assay).

AF from eggs infected with 72 of the 75 viruses that we attempted to grow showed hemagglutination activity (HAU), indicating viral presence. We collected pure virus from all virus preparations in amounts ranging from 20 to 6,920 μ g. Sham infection and purification of AF from control eggs did not produce enough protein for detection, and the preparation ran cleanly on a gel (10 μ l of a 100- μ l preparation) (see lane 29 in Fig. S1 in the supplemental material).

To determine the identity of individual bands on the Coomassie gel, we grew H1N1/H3N2 reassortant viruses A/HK/68 (HK), A/PR8/MCa (PR8), X31 (HK HA and neuraminidase [NA], PR8 background), J1 (HK HA, PR8 background), and P50 (PR8 HA, HK background). These viruses were run together on the same gel and immunoblotted with appropriate antibodies to determine which bands corresponded to which proteins via Coomassie blue staining (Fig. S2A and B). We used the following antibodies: mouse anti-HA1 monoclonal antibody (mAb) (CM1); mouse anti-HA2 mAb RA 5-22; anti-M1 (M2-1C6; anti-mouse antibody recognizing 9-kDa N-terminal fragment); anti-NP C-terminal rabbit polyclonal antibody 2364 (487 to 498); anti-NP HB-65; anti-NA C-terminal anti-rabbit polyclonal antibody.

To determine the effect of S162N, K163Q, and I216T mutations on glycosylation of pH1N1 HA, we made HA plasmids with different mutation combinations in the A/California/07/2009 HA background, which also included an additional D222G mutation to increase yield in eggs (Fig. 5C). Reassortant viruses with mutant HA and NA from A/California/07/2009 and internal segments from PR8 were rescued and grown in eggs. Virus was enriched from AF by centrifugation, standardized by enzyme-linked immunosorbent assay (ELISA) using a stem-reactive antibody (C179), and visualized by SDS-PAGE using the anti-HA CM-1 antibody. A portion of each virus was treated with PNGase F.

Protein gels and immunoblotting. We mixed purified virus (1 μ g protein) with 4 \times NuPage loading buffer (Invitrogen) and boiled the mixture for 10 min at 96°C. We electrophoresed samples with a Chameleon Duo Li-Cor ladder on 4% to 12% Bis-Tris gels (Invitrogen) at 200 V for 55 min. To visualize proteins, we fixed gels for 10 min with 10 ml 10% acetic acid and 50% methanol (with shaking) at room temperature (RT). After removal of the fixative, we added 10 ml GelCode stain (Pierce), shook the resulting preparation for 30 min at RT, and then destained the gels with water overnight. For immunoblotting, we transferred proteins from gels to polyvinylidene difluoride (PVDF) membranes with an iBLOT system at the P3 setting for 7 min. We blocked membranes for 30 min at RT using StartingBlock blocking buffer (Thermo Scientific). After incubation with primary antibody or sera for 1 h at RT and five washed performed for 5 min each wash in TBST (10 mM Tris, 150 mM NaCl, 0.1% Tween 20), we added secondary antibody, repeating the washing step for the incubation.

We imaged blots and Coomassie gels on a Li-Cor Odyssey system. Band signals and molecular weights were calculated using ImageStudio software (Li-Cor). Glycoprotein molecular weights were normalized for gel distortion, commonly called “smiling,” by adjusting each glycoprotein molecular weight by the average amount that the M1 and NP molecular weight for each virus differed from the mean molecular weight of M1 and NP proteins on each gel (Fig. S2C). Amino acid mutations in HA can shift the electrophoretic mobility of HA regardless of the presence or absence of glycan attachment. After PNGase treatment of the viruses, the electrophoretic shift caused by the amino acid mutations was minimal in our assay compared to the larger glycan shifts (Fig. S2D and E). On the basis of this finding, and to avoid overcorrection, we chose to calculate mobility shifts normalized only to NP and M proteins within the same blot. Linear regression, Pearson coefficient, and *P* values were calculated using PRISM software.

Glycan outliers. Six data points in Fig. 1 represent glycan outliers. The first, corresponding to A/Georgia/M5081/2012 (strain 28), a pH1N1 strain, is predicted to have 1 head glycan but migrated much more slowly than expected. We made a new viral stock and subjected it to deep sequencing to confirm

its identity. There were no additional predicted N-linked glycans. Among those in our panel of H1N1s (which included another contemporary pH1N1 A/California/07/2009 strain, namely, strain 27), this HA alone did not immunoblot with the anti-H1 HA monoclonal CM1 antibody even though the CM-1 peptide epitope used to generate the MAb was present. We suspect that the conformation of this protein is somehow altered under SDS-PAGE conditions, changing the HA migration through the gel and altering its interaction with CM-1. These interesting and perplexing findings will be pursued in future experiments.

Four H3N2 sequences, detailed in Table S1, have mismatched numbers of glycans compared to the other members of their clade shown in Fig. 1 and Fig. S3. Strains predicted to have more glycans than other closely related strains were found to run at levels equal to or higher than those seen with other members without the additional glycan. This indicates that even though they are predicted to have more glycans, they actually appear to be less extensively glycosylated. One of these, A/Victoria/361/2011, is the only H3N2 in our panel that is predicted to have 8 head glycans. As indicated above, H3N2s with more than 7 predicted glycans are uncommon, implying constraints on H3 HA function that lead to incomplete glycosylation during biogenesis. In contrast, the single strain with fewer predicted glycans than the other members of its group is A/Nanchang/933/1995; it migrated as fast as other H3 strains with 3 glycans.

The H1N1 A/Texas/36-JY2/1991 strain has 5 predicted glycans but migrates faster than other 5-glycan H1N1s. Strain A/Texas/36-JY2/1991 is the only representative of viruses circulating between 1987 and 1997. There are 4 glycan differences between these strains and those with A/Fort Leonard Wood/1951-type glycosylation (Fig. S3). This could account for the decreased mobility, although more sequences would be needed to support this conjecture.

Finally, three strains (Fig. 1B, open triangles) were not included in calculating glycan addition over time. Strain A/Melbourne/1/1946 was removed because its HA is most closely related to those of strains circulating 10+ years earlier (Fig. S3A). Strains A/Denver/JY2/1957 and A/Malaysia/JY2/1954 were removed because they occurred after strain A/Fort Leonard Wood/1951, the strain most closely related to strains reintroduced after 1977.

Influenza sequences and N-glycosylation prediction. Human HA protein sequences were retrieved from the NIAID Influenza Research Database (IRD) (44) through the Web site at <http://www.fludb.org> (accessed 20 April 2017). Sequences with identical strain names were retained only if the majority of members had identical sequences. Sequences were aligned to A/California/04/2009 (H1N1) (FJ966082) with MAFFT v7.305b and were removed if indels existed in relation to the reference or if any ambiguities existed within the HA head. The amino acid numbering of sites important for this work is shown (with both H1 and H3 numbering conventions) in Table S2 in the supplemental material. Potential glycosylation sites were identified by searching for the NX[S/T]Y motif, where X and Y represent any amino acid other than proline, or by using NetNGlyc 1.0 artificial neural network-based prediction software (28). The glycan frequencies and counts and the sequences used for Fig. 2 and 5 and Fig. S4 can be found in Files S2 to S10 at http://downloads.misms.net/Publications/Glycosylation_mBio/index.html. The stacked glycan plots (Fig. 2A and Fig. 3A) are similar to those reported in a previous publication (9). For the quantitative analysis of the SDS-PAGE data presented in Fig. 1 and Fig. S3, residues with positive NetNGlyc scores were considered glycosylated ("N⁺"), and residues with only one negative mark were recorded as "N⁻." Both conditions were treated as having a glycan present. Glycan predictions recorded as N⁻ were counted as negative.

Determining the duration of glycan eras is complicated by the difficulties encountered in defining criteria to delineate limits representing a starting year and ending year, which were not always sharp. We chose to measure the durations of the glycan eras using three different thresholds, namely, 10%, 50%, and 80% of sequences having the new glycan arrangement.

Phylogenetic analysis of pH1N1 viruses. To investigate the evolution and global sweep of N162/I216T glycosylated viruses, all 34,029 pH1N1 viruses available in the in GISAID EpiFlu database from 2009 to 2017 were downloaded (HA segment) and annotated with country and date of collection. The data set was aligned using MAFFT v7.310 (45). Each HA sequence was annotated for the presence of the glycosylation substitution N162/I216T and the antigenic mutation S84N. In total, 8,069 HA sequences had N162, 7,725 HA sequences had N162/I216T, and 9,538 had S84N. The 4,911 sequences representing the S84N dominant clade in File S12 (http://downloads.misms.net/Publications/Glycosylation_mBio/index.html) included 4,902 sequences with N84, 5 sequences with T84, 1 sequence with X84, and 1 sequence with S84, the latter likely representing reversion events. The 4,466 sequences representing the N162 dominant clade in File S13 (http://downloads.misms.net/Publications/Glycosylation_mBio/index.html) included 4,441 sequences with N162. Among these, 4,413 sequences were N162/I216T, 17 sequences had N162/T216I, 4 sequences had N162/T216A, 4 sequences had N162/T216P, 2 sequences had N162/T216K, and 1 sequence had N162/T216X, the latter likely representing reversion events. We estimated the proportion of pH1N1 viruses with the S84N antigenic mutation or N162/I216T glycosylation circulating in 12 regions over time, using the R program (46). Additionally, we applied a locally weighted scatterplot smoothing (LOWESS) algorithm to better visualize the fast increase of both proportions. For these analyses, we grouped countries into the regions defined in Table S3C.

We inferred a phylogenetic tree for all sequences (7,382 sequences collected from 2014 to 2017) using the maximum likelihood (ML) method available in the program RAxML v7.2.6 (47), incorporating a general time-reversible (GTR) model of nucleotide substitution with gamma-distributed (Γ) rate variation among sites. To assess the robustness of each node, a bootstrap resampling process was performed (500 replicates), again using the ML method available in RAxML v7.2.6. We identified the dominant clades with N162/I216T and S84N sequences that were first detected in 2015 and 2014, respectively, to reconstruct their early evolutionary and dispersal dynamics using a Bayesian approach.

For computational efficiency, we removed identical sequences and those shorter than 75% of the 1698 nucleotides of the HA segment, and we separately collected all N162/I216T sequences within the first year of detection (4 January 2015 to 4 January 2016, $n = 556$ sequences) and all S84N sequences within the first year and a half of detection (5 February 2014-02-05 to 22 September 2015, $n = 291$ sequences; Fig. S6B and Fig. S7B). To obtain a better geographic representation of samples, we subsampled regions with a large number of sequences, taking into account maximum time span and geographic diversity (final N162/I216T data set = 159 sequences; S84N data set = 125 sequences) (Fig. S6A and Fig. S7C).

Phylogenetic relationships were inferred by the time-scaled Bayesian approach using Markov chain Monte Carlo (MCMC), available via the BEAST v1.8.4 package (48), and the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD. A relaxed molecular clock was used, with a Hasegawa-Kishino-Yano 85 (HKY85) model (49) of nucleotide substitution with gamma-distributed rate variation among sites and either a constant population size or exponential-growth coalescent model. For viruses for which only the year of viral collection was available, the lack of tip date precision was accommodated by sampling uniformly across a 1-year window from January 1st to December 31st. The MCMC tool was run separately at least three times for each of the data sets and for at least 100 million iterations, with subsampling performed every 10,000 iterations, using the BEAGLE library to improve computational performance (50). All parameters reached convergence, as assessed visually using Tracer v1.6, with statistical uncertainty reflected in values corresponding to the 95% highest posterior density (HPD). We performed a model comparison analysis by computing the marginal likelihoods of the parametrizations with a constant population size or with the exponential-growth coalescent models, measured by the use of the harmonic mean estimator and the Akaike information criteria through MCMC (AICM). At least 10% of the chain was removed as burn-in, and parallel runs were combined using LogCombiner v1.8.4 and downsampled to generate a final posterior distribution of 500 trees that was used in the subsequent spatial analysis (51). We found that for both data sets, the exponential-growth model performed better, providing a lower AICM value and positive harmonic mean values that indicate better relative model fit (Table S3B). Results reported above show this parametrization. Nonetheless, the data representing the divergence times are robust, as the highest posterior density intervals largely overlap for both parametrizations (Fig. S3B).

The phylogeographic analysis considered 11 and 12 locations for the S84N and N162/I216T data sets, respectively (Table S3A). A nonreversible discrete model was used to infer the rate of location transitions, along with Bayesian stochastic search variable selection (BSSVS) to identify those highly significant transitions while improving statistical efficiency (52). For computational efficiency, the phylogeographic analysis was run using an empirical distribution of 500 trees (51), allowing the MCMC chain to be run for 50 million iterations, with sampling every 5,000 iterations. Maximum clade credibility (MCC) trees were summarized using TreeAnnotator v1.8.4, and the trees were visualized in FigTree v1.4.3.

HA homology model. We created a homology model of the influenza A virus HA head domain with Schrödinger's Prime module, using the sequence of the Shandong 2009 H1N1 strain (UniProt accession number F2YI86) and the 2WRO structure as a template for the homotrimer assembly. To model the HA stalk, the 2WPQ structure was identified as a good candidate for trimeric alpha-helical bundles.

Data availability. The reagents and all other raw data are available from J. W. Yewdell upon request. Processed data and sequences supporting this study can be found at http://downloads.misms.net/Publications/Glycosylation_mBio/index.html.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mBio.00204-19>.

FIG S1, TIF file, 2.2 MB.

FIG S2, TIF file, 2.3 MB.

FIG S3, TIF file, 2.3 MB.

FIG S4, TIF file, 1.1 MB.

FIG S5, TIF file, 1.2 MB.

FIG S6, TIF file, 1.5 MB.

FIG S7, TIF file, 1 MB.

TABLE S1, PDF file, 0.05 MB.

TABLE S2, PDF file, 0.04 MB.

TABLE S3, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

This work was supported by the Division of Intramural Research, NIAID. Glennys Reynoso provided outstanding technical support. Jia Jie Wei, Mina Seedhom, Davide Angeletti, and Pascal Gagneux provided insightful conversations. Zhiping Ye at the FDA generously supplied viruses.

S.E.H. is funded through NIAID grants 1R01AI113047 and 1R01AI108686. N.S.T. is supported by NIAID's Centers of Excellence in Influenza Virus Research and Surveillance (HHSN27772201400008C).

I.K. and M.O.A. grew, purified, and analyzed viruses. M.A., N.S.T., M.I.N., and M.O.A. developed and performed bioinformatic analysis. S.E.H., S.J.Z., and J.S.G. rescued and analyzed glycan mutants. R.E.A. and L.C. performed the HA modeling. N.S.T., M.I.N., M.A., S.J.Z., L.C., and M.O.A. developed the figures. J.W.Y. and M.O.A. designed and supervised the study and then wrote the manuscript with input from the rest of us.

REFERENCES

- WHO. 2003. Influenza fact sheet no. 211. <http://www.who.int/media/centre/factsheets/2003/fs211/en/>. Accessed 21 June 2018.
- Couch RB, Kasel JA. 1983. Immunity to influenza in man. *Annu Rev Microbiol* 37:529–549. <https://doi.org/10.1146/annurev.mi.37.100183.002525>.
- CDC. 2017. Vaccine effectiveness-how well does the flu vaccine work? <https://www.cdc.gov/flu/about/qa/vaccineeffect.htm>. Accessed 21 June 2018.
- Keil W, Geyer R, Dabrowski J, Dabrowski U, Niemann H, Stirn S, Klenk HD. 1985. Carbohydrates of influenza virus. Structural elucidation of the individual glycans of the FPV hemagglutinin by two-dimensional ¹H n.m.r. and methylation analysis. *EMBO J* 4:2711–2720. <https://doi.org/10.1002/j.1460-2075.1985.tb03991.x>.
- Khatiri K, Klein JA, White MR, Grant OC, Leymarie N, Woods RJ, Hartshorn KL, Zaia J. 2016. Integrated omics and computational glycobiology reveal structural basis for influenza A virus glycan microheterogeneity and host interactions. *Mol Cell Proteomics* 15:1895–1912. <https://doi.org/10.1074/mcp.M116.058016>.
- Daniels R, Kurowski B, Johnson AE, Hebert DN. 2003. N-linked glycans direct the cotranslational folding pathway of influenza hemagglutinin. *Mol Cell* 11:79–90. [https://doi.org/10.1016/S1097-2765\(02\)00821-3](https://doi.org/10.1016/S1097-2765(02)00821-3).
- Zhang M, Gaschen B, Blay W, Foley B, Haigwood N, Kuiken C, Korber B. 2004. Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin. *Glycobiology* 14:1229–1246. <https://doi.org/10.1093/glycob/cwh106>.
- Sun S, Wang Q, Zhao F, Chen W, Li Z. 2011. Glycosylation site alteration in the evolution of influenza A (H1N1) viruses. *PLoS One* 6:e22844. <https://doi.org/10.1371/journal.pone.0022844>.
- Lee PS, Ohshima N, Stanfield RL, Yu W, Iba Y, Okuno Y, Kurosawa Y, Wilson IA. 2014. Receptor mimicry by antibody F045-092 facilitates universal binding to the H3 subtype of influenza virus. *Nat Commun* 5:3614. <https://doi.org/10.1038/ncomms4614>.
- An Y, McCullers JA, Alymova I, Parsons LM, Cipollo JF. 2015. Glycosylation analysis of engineered H3N2 influenza A virus hemagglutinins with sequentially added historically relevant glycosylation sites. *J Proteome Res* 14:3957–3969. <https://doi.org/10.1021/acs.jproteome.5b00416>.
- Su YC, Bahl J, Joseph U, Butt KM, Peck HA, Koay ES, Oon LL, Barr IG, Vijaykrishna D, Smith GJ. 2015. Phylodynamics of H1N1/2009 influenza reveals the transition from host adaptation to immune-driven selection. *Nat Commun* 6:7952. <https://doi.org/10.1038/ncomms8952>.
- Sun X, Jayaraman A, Mani Prasad P, Raman R, Houser KV, Pappas C, Zeng H, Sasisekharan R, Katz JM, Tumpey TM. 2013. N-linked glycosylation of the hemagglutinin protein influences virulence and antigenicity of the 1918 pandemic and seasonal H1N1 influenza A viruses. *J Virol* 87:8756–8766. <https://doi.org/10.1128/JVI.00593-13>.
- Medina RA, Stertz S, Manicassamy B, Zimmermann P, Sun X, Albrecht RA, Uusi-Kerttula H, Zagordi O, Belshe RB, Frey SE, Tumpey TM, García-Sastre A. 2013. Glycosylations in the globular head of the hemagglutinin protein modulate the virulence and antigenic properties of the H1N1 influenza viruses. *Sci Transl Med* 5:187ra70. <https://doi.org/10.1126/scitranslmed.3005996>.
- Job ER, Deng YM, Barfod KK, Tate MD, Caldwell N, Reddiex S, Maurer-Stroh S, Brooks AG, Reading PC. 2013. Addition of glycosylation to influenza A virus hemagglutinin modulates antibody-mediated recognition of H1N1 2009 pandemic viruses. *J Immunol* 190:2169–2177. <https://doi.org/10.4049/jimmunol.1202433>.
- Vigerust DJ, Ulett KB, Boyd KL, Madsen J, Hawgood S, McCullers JA. 2007. N-linked glycosylation attenuates H3N2 influenza viruses. *J Virol* 81:8593–8600. <https://doi.org/10.1128/JVI.00769-07>.
- Wu NC, Zost SJ, Thompson AJ, Oyen D, Nycholat CM, McBride R, Paulson JC, Hensley SE, Wilson IA. 2017. A structural explanation for the low effectiveness of the seasonal influenza H3N2 vaccine. *PLoS Pathog* 13:e1006682. <https://doi.org/10.1371/journal.ppat.1006682>.
- Fonville JM, Wilks SH, James SL, Fox A, Ventresca M, Aban M, Xue L, Jones TC, Le NMH, Pham QT, Tran ND, Wong Y, Mosterin A, Katzelnick LC, Labonte D, Le TT, van der Net G, Skepner E, Russell CA, Kaplan TD, Rimmelzwaan GF, Masurel N, de Jong JC, Palache A, Beyer WEP, Le QM, Nguyen TH, Wertheim HFL, Hurt AC, Osterhaus A, Barr IG, Fouchier RAM, Horby PW, Smith DJ. 2014. Antibody landscapes after influenza virus infection or vaccination. *Science* 346:996–1000. <https://doi.org/10.1126/science.1256427>.
- Aytay S, Schulze IT. 1991. Single amino acid substitutions in the hemagglutinin can alter the host range and receptor binding properties of H1 strains of influenza A virus. *J Virol* 65:3022–3028.
- Das SR, Hensley SE, David A, Schmidt L, Gibbs JS, Puigbo P, Ince WL, Binnink JR, Yewdell JW. 2011. Fitness costs limit influenza A virus hemagglutinin glycosylation as an immune evasion strategy. *Proc Natl Acad Sci U S A* 108:E1417–E1422. <https://doi.org/10.1073/pnas.1108754108>.
- Cobey S, Hensley SE. 2017. Immune history and influenza virus susceptibility. *Curr Opin Virol* 22:105–111. <https://doi.org/10.1016/j.coviro.2016.12.004>.
- Blackburne BP, Hay AJ, Goldstein RA. 2008. Changing selective pressure during antigenic changes in human influenza H3. *PLoS Pathog* 4:e1000058. <https://doi.org/10.1371/journal.ppat.1000058>.
- Sun H, Yang J, Zhang T, Long L-P, Jia K, Yang G, Webby RJ, Wan X-F. 2013. Using sequence data to infer the antigenicity of influenza virus. *mBio* 4:e00230-13. <https://doi.org/10.1128/mBio.00230-13>.
- Alymova IV, York IA, Air GM, Cipollo JF, Gulati S, Baranovich T, Kumar A, Zeng H, Ganseboom S, McCullers JA. 2016. Glycosylation changes in the globular head of H3N2 influenza hemagglutinin modulate receptor binding without affecting virus virulence. *Sci Rep* 6:36216. <https://doi.org/10.1038/srep36216>.
- Yang H, Carney PJ, Chang JC, Guo Z, Villanueva JM, Stevens J. 2015. Structure and receptor binding preferences of recombinant human A(H3N2) virus hemagglutinins. *Virology* 477:18–31. <https://doi.org/10.1016/j.virol.2014.12.024>.
- de Graaf M, Fouchier RA. 2014. Role of receptor binding specificity in influenza A virus transmission and pathogenesis. *EMBO J* 33:823–841. <https://doi.org/10.1002/emboj.201387442>.
- Park S, Lee I, Kim JI, Bae JY, Yoo K, Kim J, Nam M, Park M, Yun SH, Cho WI, Kim YS, Ko YY, Park MS. 2016. Effects of HA and NA glycosylation pattern changes on the transmission of avian influenza A(H7N9) virus in guinea pigs. *Biochem Biophys Res Commun* 479:192–197. <https://doi.org/10.1016/j.bbrc.2016.09.024>.
- Hrinčič ER, Liedmann S, Finkelstein D, Vogel P, Ganseboom S, Samarasinghe AE, You D, Cormier SA, McCullers JA. 2015. Acute lung injury results from innate sensing of viruses by an ER stress pathway. *Cell Rep* 11:1591–1603. <https://doi.org/10.1016/j.celrep.2015.05.012>.
- Gupta RBS. 2004. Prediction of N-glycosylation sites in human proteins. <http://www.cbs.dtu.dk/services/NetNGlyc/>.
- Tsuchiya E, Sugawara K, Hongo S, Matsuzaki Y, Muraki Y, Li ZN, Nakamura K. 2001. Antigenic structure of the haemagglutinin of human influenza A/H2N2 virus. *J Gen Virol* 82:2475–2484. <https://doi.org/10.1099/0022-1317-82-10-2475>.
- Bauc E. 1984. Model studies on N-glycosylation of proteins. *Biochem Soc Trans* 12:514–517. <https://doi.org/10.1042/bst0120514>.
- Reddy A, Gibbs BS, Liu YL, Coward JK, Changchien LM, Maley F. 1999. Glycosylation of the overlapping sequons in yeast external invertase: effect of amino acid variation on site selectivity in vivo and in vitro. *Glycobiology* 9:547–555. <https://doi.org/10.1093/glycob/9.6.547>.
- Shrimal S, Gilmore R. 2013. Glycosylation of closely spaced acceptor sites in human glycoproteins. *J Cell Sci* 126:5513–5523. <https://doi.org/10.1242/jcs.139584>.

33. Yin Y, Zhang X, Qiao Y, Wang X, Su Y, Chen S, Qin T, Peng D, Liu X. 2017. Glycosylation at 11Asn on hemagglutinin of H5N1 influenza virus contributes to its biological characteristics. *Vet Res* 48:81. <https://doi.org/10.1186/s13567-017-0484-8>.
34. Liu YJ, Wu SL, Love KR, Hancock WS. 2017. Characterization of site-specific glycosylation in influenza A virus hemagglutinin produced by *Spodoptera frugiperda* insect cell line. *Anal Chem* 89:11036–11043. <https://doi.org/10.1021/acs.analchem.7b03025>.
35. Gerhard W, Yewdell J, Frankel ME, Webster R. 1981. Antigenic structure of influenza virus haemagglutinin defined by hybridoma antibodies. *Nature* 290:713–717. <https://doi.org/10.1038/290713a0>.
36. Linderman SL, Chambers BS, Zost SJ, Parkhouse K, Li Y, Herrmann C, Ellebedy AH, Carter DM, Andrews SF, Zheng NY, Huang M, Huang Y, Strauss D, Shaz BH, Hodinka RL, Reyes-Teran G, Ross TM, Wilson PC, Ahmed R, Bloom JD, Hensley SE. 2014. Potential antigenic explanation for atypical H1N1 infections among middle-aged adults during the 2013–2014 influenza season. *Proc Natl Acad Sci U S A* 111:15798–15803. <https://doi.org/10.1073/pnas.1409171111>.
37. Huang KY, Rijal P, Schimanski L, Powell TJ, Lin TY, McCauley JW, Daniels RS, Townsend AR. 2015. Focused antibody response to influenza linked to antigenic drift. *J Clin Invest* 125:2631–2645. <https://doi.org/10.1172/JCI81104>.
38. Ben-Dor S, Esterman N, Rubin E, Sharon N. 2004. Biases and complex patterns in the residues flanking protein N-glycosylation sites. *Glycobiology* 14:95–101. <https://doi.org/10.1093/glycob/cwh004>.
39. Eggink D, Goff PH, Palese P. 2014. Guiding the immune response against influenza virus hemagglutinin toward the conserved stalk domain by hyperglycosylation of the globular head domain. *J Virol* 88:699–704. <https://doi.org/10.1128/JVI.02608-13>.
40. Tate MD, Job ER, Deng YM, Gunalan V, Maurer-Stroh S, Reading PC. 2014. Playing hide and seek: how glycosylation of the influenza virus hemagglutinin can modulate the immune response to infection. *Viruses* 6:1294–1316. <https://doi.org/10.3390/v6031294>.
41. Brooke CB. 2017. Population diversity and collective interactions during influenza virus infection. *J Virol* 91:e01164-17.
42. Parra GI, Squires RB, Karangwa CK, Johnson JA, Lepore CJ, Sosnovtsev SV, Green KY. 2017. Static and evolving norovirus genotypes: implications for epidemiology and immunity. *PLoS Pathog* 13:e1006136. <https://doi.org/10.1371/journal.ppat.1006136>.
43. Altman MO, Bennink JR, Yewdell JW, Herrin BR. 2015. Lamprey VLRB response to influenza virus supports universal rules of immunogenicity and antigenicity. *Elife* 4:e07467. <https://doi.org/10.7554/eLife.07467>.
44. Zhang Y, Aevermann BD, Anderson TK, Burke DF, Dauphin G, Gu Z, He S, Kumar S, Larsen CN, Lee AJ, Li X, Macken C, Mahaffey C, Pickett BE, Reardon B, Smith T, Stewart L, Suloway C, Sun G, Tong L, Vincent AL, Walters B, Zaremba S, Zhao H, Zhou L, Zmasek C, Klem EB, Scheuermann RH. 2017. Influenza Research Database: an integrated bioinformatics resource for influenza virus research. *Nucleic Acids Res* 45:D466–D474. <https://doi.org/10.1093/nar/gkw857>.
45. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>.
46. Team RC. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.R-project.org>.
47. Stamatakis A. 2006. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690. <https://doi.org/10.1093/bioinformatics/btl446>.
48. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29:1969–1973. <https://doi.org/10.1093/molbev/mss075>.
49. Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 22:160–174. <https://doi.org/10.1007/BF02101694>.
50. Suchard MA, Rambaut A. 2009. Many-core algorithms for statistical phylogenetics. *Bioinformatics* 25:1370–1376. <https://doi.org/10.1093/bioinformatics/btp244>.
51. Pagel M, Meade A, Barker D. 2004. Bayesian estimation of ancestral character states on phylogenies. *Syst Biol* 53:673–684. <https://doi.org/10.1080/10635150490522232>.
52. Lemey P, Rambaut A, Drummond AJ, Suchard MA. 2009. Bayesian phylogeography finds its roots. *PLoS Comput Biol* 5:e1000520. <https://doi.org/10.1371/journal.pcbi.1000520>.
53. Neher RA, Bedford T. 2015. nextflu: real-time tracking of seasonal influenza virus evolution in humans. *Bioinformatics* 31:3546–3548. <https://doi.org/10.1093/bioinformatics/btv381>.
54. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T, Neher RA. 2018. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 34:4121–4123. <https://doi.org/10.1093/bioinformatics/bty407>.
55. FluNet W. 2018. http://www.who.int/influenza/gisrs_laboratory/flunet/en/.
56. Burke DF, Smith DJ. 2014. A recommended numbering scheme for influenza A HA subtypes. *PLoS One* 9:e112302. <https://doi.org/10.1371/journal.pone.0112302>.
57. Mathys L, Balzarini J. 2015. Several N-glycans on the HIV envelope glycoprotein gp120 preferentially locate near disulphide bridges and are required for efficient infectivity and virus transmission. *PLoS ONE* 10:e0130621. <https://doi.org/10.1371/journal.pone.0130621>.