Hemagglutinin Glycan Clock Guides Human Influenza A Virus Evolution

Meghan O. Altman[1]*, Matthew Angel[1], Ivan Košík[1], James S. Gibbs[1], Nídia S. Trovão[2,3], Martha I. Nelson[2], Jonathan W. Yewdell[1]*

[1] Cellular Biology Section, Laboratory of Viral Diseases NIAID, NIH, Bethesda MD 20892

[2] Division of International Epidemiology and Population Studies, Fogarty International Center, NIH, Bethesda MD 20892

[3] Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, NY 11766

* Corresponding authors

ABSTRACT

Glycosylation of influenza A virus hemagglutinin (HA) globular domain greatly influences viral antigenicity, receptor binding, and innate immunity. Gel electrophoresis and bioinformatics analysis reveals that circulating human H1 and H3 HA glycans largely evolve in a clock-like progression according to three rules. This glycan clock predicted the 2015-2016 selective sweep of pH1N1 by strains with an added glycan at residue 179. A limit on glycan shielding of the HA head dictates the ultimate fate of circulating strains, enabling our prediction that circulating H3 HA strains will either undergo a glycan readjustment after 9 years or be supplanted by a novel pandemic virus.

INTRODUCTION

Seasonal influenza A virus (IAV) sickens millions and kills hundreds of thousands of people annually (WHO, 2003). IAV remains endemic despite high levels of adaptive immunity in virtually all humans after early childhood. IAV persistence is enabled by antigenic drift: largely due to the rapid selection of viruses with amino acid substitutions in the hemagglutinin (HA) glycoprotein that enable viral escape from HA-specific antibody-based immunity. HA is a homotrimeric glycoprotein with a globular head domain resting atop a stalk 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 infection. Antibody (Ab) binding to the head neutralizes virus by blocking attachment or conformational alterations required for HA-mediated membrane fusion. Understanding Ab-driven HA evolution is essential to improving influenza vaccination, which at best, offers limited protection from infection (CDC, 2017; Couch and Kasel, 1983).

As newly synthesized HA enters the endoplasmic reticulum of IAV infected cells, oligosaccharides are attached to Asn residues present in the motif Asn-X-Ser/Thr-Y, where X/Y are any amino acids except Pro. Attached glycans can be high-mannose glycans or larger more complex branched glycans (Keil et al., 1985; Khatri et al., 2016). HAs possess multiple highly conserved glycans on the stem domain crucial for HA folding and oligomerization (Daniels et al., 2003). Glycosylation of the head is much more variable, as evolving HAs add, retain, and occasionally lose potential head glycosylation sites (An et al., 2015; Sun et al., 2011; Vijaykrishna et al., 2015). Head glycans can promote viral fitness by shielding virus from antibody binding and tuning receptor affinity and specificity (Job et al., 2013; Medina et al., 2013; Sun et al., 2013b; Vigerust et al., 2007). Conversely, glycans can deleteriously modulate receptor binding, enhance viral neutralization by innate immune lectins, or increase stress in the endoplasmic reticulum during translation (Alymova et al., 2016; Das et al., 2011; de Graaf and Fouchier, 2014; Hrincius et al., 2015; Park et al., 2016; Yang et al., 2015). Glycosylation changes are generally not correlated with changes between antigenic clusters, in part, because glycan addition typically lowers HA receptor avidity (Aytay and Schulze, 1991), complicating binary assays such as hemagglutination inhibition (Das et al., 2011). Consequently, they are typically unaccounted for in molecular evolution studies using HI-based analytic tools such as antigenic cartography (Blackburne et al., 2008; Sun et al., 2013a).

Quantitative measurements of HA suggest rules for glycan evolution

Studies of HA glycosylation evolution rely nearly exclusively on bioinformatic prediction of glycan addition. NetNGlyc, the state of the art algorithm (Gupta, 2004), imperfectly predicts N-linked glycosylation (Khatri et al., 2016). We therefore used gel electrophoresis to monitor glycan addition to HAs from 72 egg-grown H1N1 and H3N2 strains based on decreased HA migration (**Fig. 1**, **fig. S1-S4**). HA migration strongly correlates with the number of computationally predicted head glycosylation sites (Pearson coefficient r = 0.96, $P < 0.01$) (**Fig. 1A**). H1 adds an apparent 3.9 ± 0.1 kDa [$R^2 = 0.89$] per glycan, suggesting the majority of such glycans are highly branched. The average H3 glycan, by contrast, is smaller (3.0 ± 0.2 kDa [$R^2 = 0.90$]), suggesting heterogeneity in glycan mass or incomplete glycosylation. These findings are consistent with mass spectrometry characterization of glycans from representative viruses showing that H1 glycans skew to larger more complex structures, while H3 glycans at several sites are majority high-mannose (Khatri et al., 2016; Sun et al., 2013b).

Importantly, for H1 and H3 HAs, the total glycan mass added to the head during evolution in humans is, respectively, 15.5 kDa and 15 kDa. This is consistent with the idea of an upper limit of glycan shielding for H1 and H3 HA head domains totaling to ~20 kDa (5 glycans for H1, 7 for H3). This is not strictly due to the replicative capacity of IAVs: H1 HAs can accommodate at least 8 head glycans while maintaining *in vitro* fitness (Eggink

et al., 2014). Rather it points to glycan-based fitness costs *in vivo*, possibly due innate immune mechanisms (Tate et al., 2014).

Plotting HA migration versus time reveals the history of seasonal H1N1 (sH1N1) glycosylation (**Fig. 1B**, triangles). First-wave sH1N1 (1933-1957) HAs behaved differently than the second-wave sH1N1 representatives (1977-2008). During the first-wave, HA increased by an average of 3.9 kDa (or one glycan) every 5 years (slope = 0.69 kDa/year, R^2 = 0.72). The H1N1 strains that re-emerged after the 1977 introduction of H1N1 exhibited similar HA migration to the 5-head glycan A/Fort Leonard Wood/1951 (inside blue box, **Fig. 1B**). Rather than increasing in number, electrophoresis confirms the NetNGlyc prediction that head glycans were maintained until decreasing by one unit after 1991.

Fig. 1

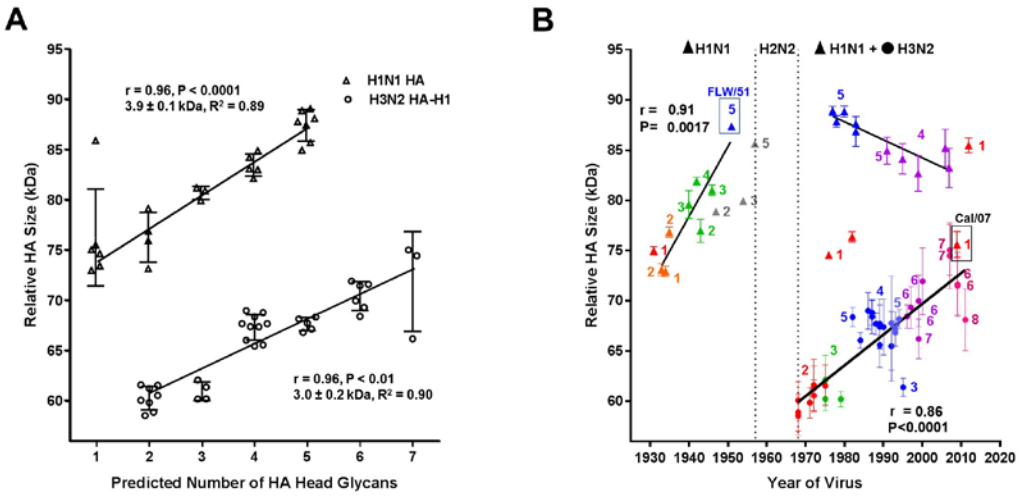


Fig. 1. Quantitative survey of glycan evolution in IAV. (A) Relative HA size, measured by migration rate through a gel, correlates with predicted number of head glycans. H1N1s (triangles) add 3.9 ± 0.1 kD per glycan, and 15.5 kD total between 1 and 5 glycans. H3N2s (HA1, circles) add 3.0 ± 0.2 kD per glycan and 15 kD total between 2 and 7 glycans. **(B)** HA size over time for H1N1s (triangles) and H3N2s (HA1, circles). Strains are color-coded to match glycan groups in **fig. S2.** Numbers of head glycans per strain are noted next to individual points. A glycan is added on average every 5 years for sH1N1 (3.9 kD) and every 8 years for H3N2 (3.0 kD). The strain most related to sH1N1 reintroduced in 1977, A/Fort Leonard Wood/1951, is boxed and labelled in blue. The pH1N1 strain, A/California/07/2009, is boxed and labelled in black. Pearson coefficient of correlation (r) and its P value are shown on both graphs. Error bars are standard error of the means from n = 3-5 blots.

Bioinformatic analysis supports rules of glycan evolution

The second-wave sH1N1 began in 1977 with the re-introduction of a strain genetically similar to A/Fort Leonard Wood/1951, possibly due to a flawed vaccine trial (Rozo and Gronvall, 2015) (**Fig. 2**). With 5 head glycans, reemerged sH1N1 reached our proposed H1 head-glycan ceiling compatible with competitive human circulation. Indeed, sequences predicted to have 6 glycans are unusual over the century of H1 evolution (1918-2017), except in 1986 and 1987, when they competed among H1 strains (**fig S5A**). Rather than adding glycans, reemerged sH1N1 viruses exhibit glycan adjustments at roughly 10 year intervals. In 1987, N71 pH1N1 (numbering convention is shown in Table S1) replaces N172, as N144 toggles to N142, which though adjacent to 144, is oriented to shield different epitopes. In 1997, N286 is lost, and this glycan configuration remained constant until sH1N1 became extinct in 2009.

Turning to H3 glycan evolution (**Fig. 4**), original (1968) pandemic H3 strains have two head glycans. Six years after the pandemic, H3N2 added a third glycan, N126. Six years later, it added another glycan at either N246 or N122. The virus then persisted 12 years with 4 glycans, suggesting a functional limitation to adding a fifth glycan. In 1993, an antigenically distinct H3N2 swept the globe; this virus added a glycan at N276, which is adjacent to a disulfide-bonded cysteine. This 5-glycan H3N2 disappears in 3 years, and by 1998 is completely replaced by HAs with 2 added glycans, N122 and N133, for a net increase of one glycan. It takes another 6 years for complete replacement by strains that have a seventh glycan, N144.

With 7 glycans, H3 HAs had increased 15 kDa from the original 1968 strain. It is apparent from the global human influenzome (GHI) that H3 HAs with more than 7 glycans are unfit for human circulation—8-glycan strains make up <1% of H3N2 sequences in any year (**fig. S5B**). Unable to add a glycan to escape immune pressure, 12 years after reaching the glycan limit, H3N2 underwent a selective sweep that toggled the glycan from N144 to N158 by 2015 (**fig. S6A**).

Influenza Glycan Clock algorithm

Based on these observations we propose 3 rules that explain, albeit imperfectly, the evolution of HA glycans:

1. Glycans are limited on the HA head to 5 and 7 glycans, respectively, for H1 and H3 HAs, with a total mass of ~20 kDa due to preferred use of complex (H1) *vs.* simple (H3) glycans.
2. If HA can add a glycan and retain high transmissibility, it will add one after 5 years for H1 or 6 years for H3.
3. If a glycan addition cannot occur, after 10 years for H1, or 12 years for H3, either a glycan switch will occur or a new pandemic strain will replace it.

Fig. 2

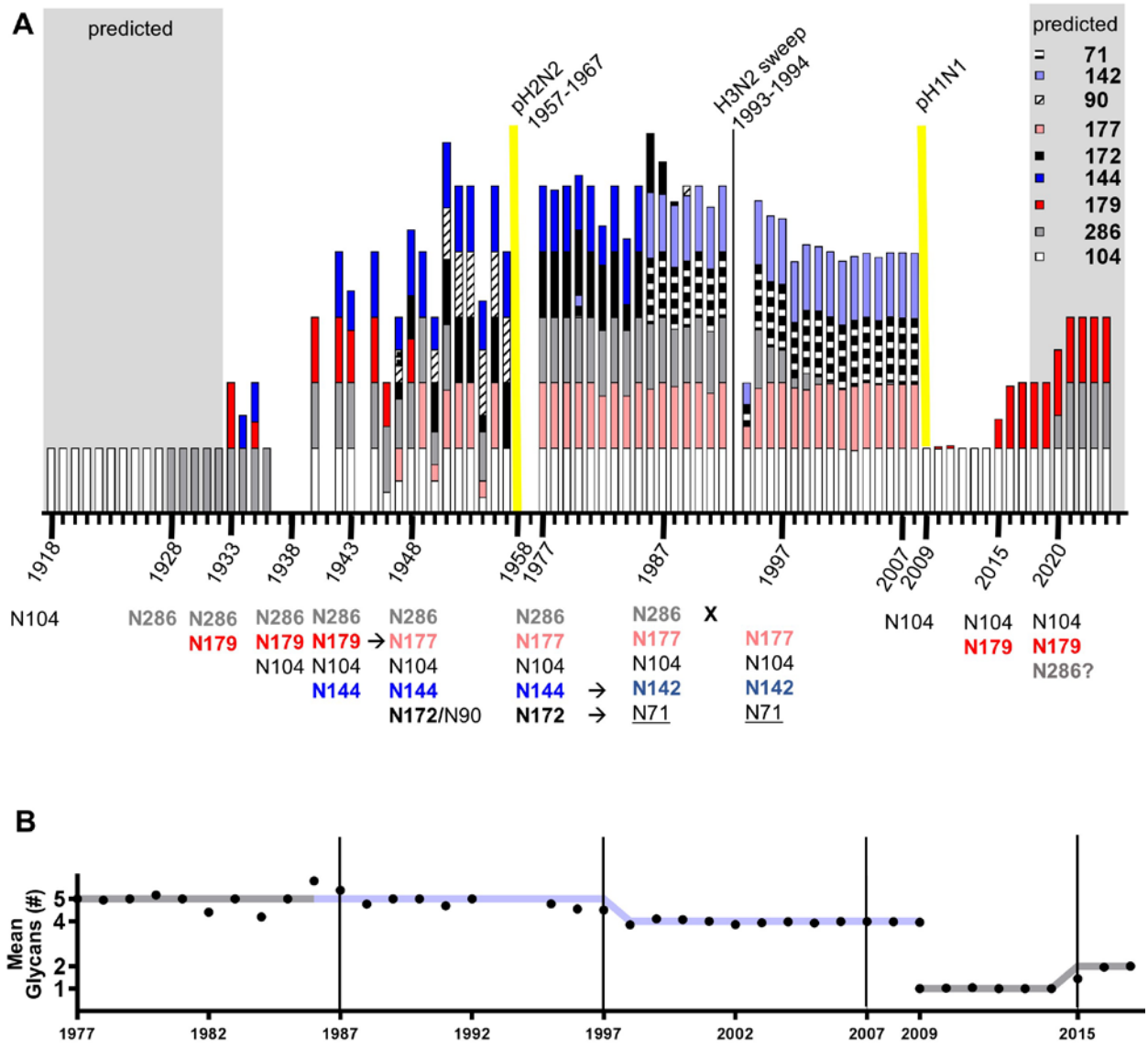


Fig. 2. H1 Glycan Clock 1918-2022. Stacked bars correspond to the percentage of all human H1N1 sequences from a given year containing a glycan listed in the legend. Dates of glycan transition are written on the timeline below along with the glycosylated residues. Projected values for strains from 1918-1933 and 2017-2022 are shown in shaded regions. Years with no data are left blank. Sequence counts for each year are shown in **fig. S6**. Pandemics are denoted by yellow lines. (B) The mean predicted number of glycans among all sequences from each year are shown as black circles. This data is laid atop the number of glycans predicted from the glycan clock algorithm (grey line, light blue line indicates glycan transition from sequences containing N144/N172 to N142/N71). Black vertical lines denote predicted years for glycan evolution.

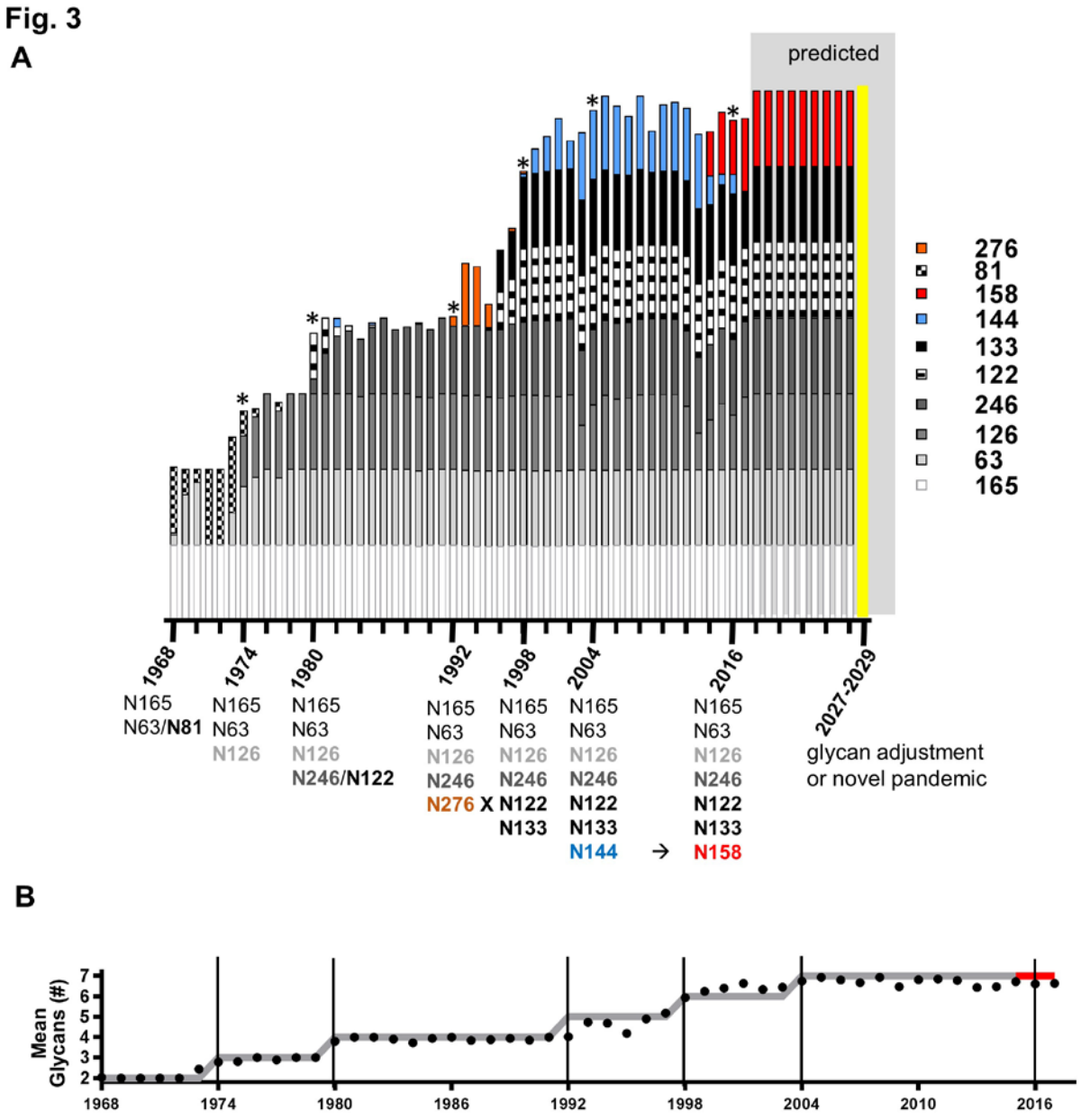


Fig. 3. H3 Glycan Clock 1968-2029. (A) Stacked bars correspond to the percentage of all human H3N2 sequences from a given year containing a glycan listed in the legend. Dates of glycan transition are written on the timeline below along with the glycosylated residues. Projected values for strains from 2017-2029 are shown in shaded regions. Potential pandemic is denoted by yellow line. **(B)** The mean predicted number of glycans among all sequences from each year are shown as black circles. This data is laid atop the number of glycans predicted from the glycan clock algorithm (grey line, red line indicates glycan transition from sequences containing N144 to N158). Black vertical lines denote years of glycan evolution.

Predicting future influenza evolution events

Using these rules, we can predict future IAV glycosylation evolution (**Fig. 2** and **3**, grey boxes). H3N2 toggled N144 to N158 over the 2014-2015 season. According to Rule 1, H3N2 is limited to 7 glycans. By Rule 3, during the 2027-2029 seasons, H3N2 should either exhibit another glycan adjustment or be replaced by a new pandemic strain.

In 2009, pandemic H1N1 (pH1N1) rapidly replaced sH1N1 in the GHI. A single head residue is predicted to be glycosylated (N104). Indeed, the HA of a representative early strain, A/California/07/2009 (**Fig. 1B**, inside black box), co-migrated with strains from the 1930s and two other swine-origin human strains, A/New Jersey/8/1976 and A/Memphis/4/1982, each with a single predicted head glycan site (**Fig. 1B**, red triangles).

Based on the addition of a globular-domain glycan to first wave H1 strains every 5 years from 1933 to 1951, we predicted that pH1N1 viruses should add a glycan by 2015. Indeed, identifying all potential HA head glycosylation sites in pH1 sequences in the Flu Database (FluDB) revealed a second predicted glycan based on a S to N change at residue 179 (**Fig. 4A**). N179 viruses are present at low frequencies in each season since pH1N1 introduction (**Fig. 4B**), but over 15 months during 2015-2016, N179 replaced all previously circulating strains in the pH1N1 GHI (**fig. S6B**).

How did the N179 strains come to selectively sweep the human pH1N1 GHI? We identified the earliest genetic precursor “sweeper strains” (SWSs) of the N179 strains in the FluDB by aligning pre-2016 N179-containing strains with ClustalΩ software to generate an evolutionary tree. The first pre-SWSs appeared almost simultaneously in January 2015 in Iran (4 strains), Turkey (1 strain), and Taiwan (1 strains). In addition to an additional glycan site at N179, all SWSs carried the I233T substitution. This combination of I233T *and* N179 is not found in prior strains.

The A/California/07/2009 HA structure provides a ready explanation for the co-selection. Residues 233 and 179 are in close proximity across the trimer interface (**Fig. 4C**), with the change from bulkier and more hydrophobic Ile to Thr likely to facilitate the addition of the large and polar glycan on the other side of the trimer interface. Notably, in H3 strains with a N179 glycan, the residue analogous to 233 in pH1N1 is also a small polar residue (Ser).

As the first 3 glycans added to sH1N1 were N104, N179, and N286 (see strain A/Hickox/JY2/1940), we predict that N286 will be present in the next glycan selective sweep strain, which based on Rule 2, should occur during the 2020-2021 season.

Fig. 4

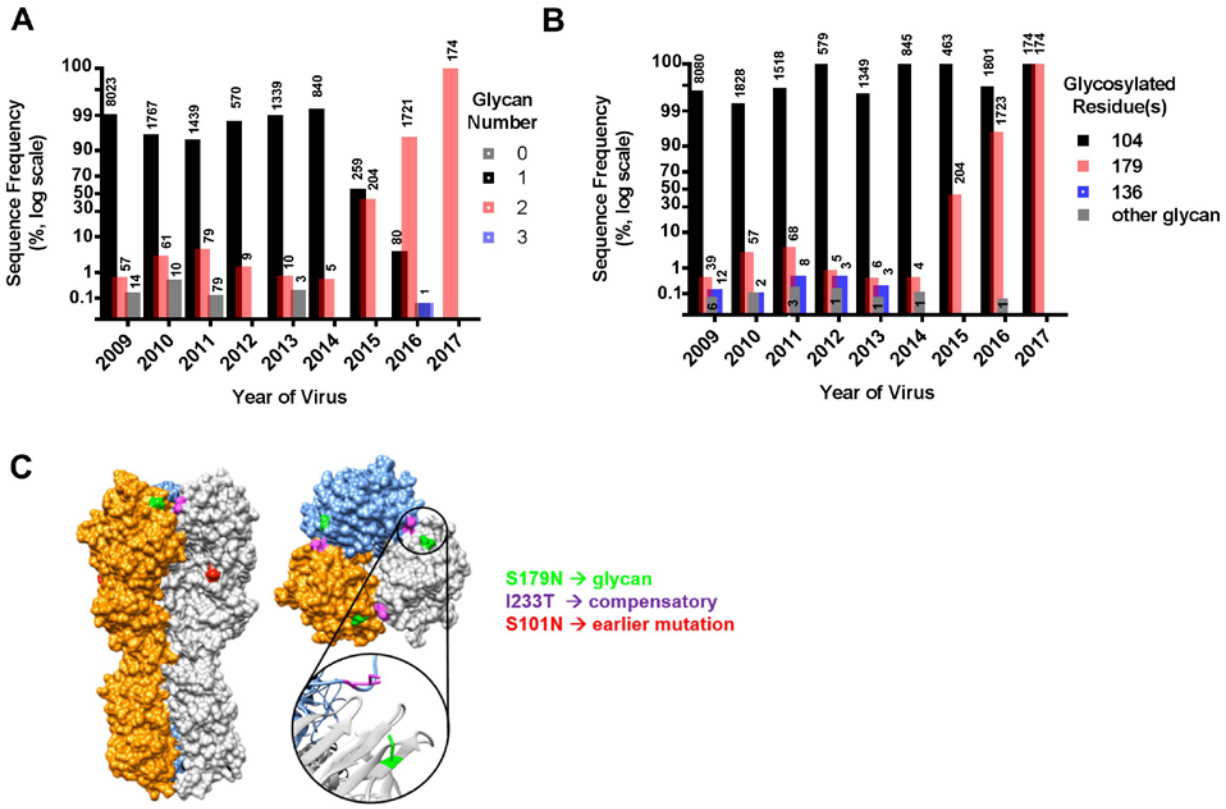


Figure 2. Glycan clock predicts recent pH1N1 glycan addition. (A) Frequency of human pH1N1 sequences in FluDB by number of predicted glycans in each calendar year. The number of sequences represented by each bar is shown on top of bar. Until January 2015 >90% of sequences had 1 glycan (black). Starting in 2015, strains with 2 glycans swept (red). **(B)** Frequency of the individual residues glycosylated in human pH1N1 sequences. N104 predominates throughout (>95%, black). N179 (red) and N136 (blue) appear at low frequencies until 2015 when strains having both N179 and N104 extinguish N104-only strains. N179 strains are found in every year. Rarely, in addition to N104 a residue other than N179 or N136 has a potential glycosylation site (grey). **(D)** A/California/07/2009 (PDB: 3LZG) hemagglutinin crystal structure with N179 highlighted in green and the compensatory mutation I233T in purple adjacent across the trimer. Inset shows magnified view of the two mutated residues.

CONCLUSION

It has been known for some time that IAV maintains circulation in humans, in part, by altering the number and type of N-linked glycans on the globular head of HA. After measuring the relative size of HA within a historical panel of IAVs we hypothesized: (1) There is a functional limit to the number of HA-head glycans IAV can add while still retaining fitness in the human population; and (2) Glycan evolution in IAV occurs at a

regular rate. When we tested these hypotheses, bioinformatically against the set of all human influenza sequences, clear patterns emerged, leading to development of the Influenza Glycan Clock algorithm. This algorithm predicted the 2015-2016 selective sweep of pH1N1 by strains with a new glycan at N179.

Going forward, we envision this algorithm will permit us to predict the timeline and trajectory of IAV evolution from their pandemic introduction to eventual replacement. We anticipate this discovery will have important implications for improving pandemic preparation and vaccine development.

ACKNOWLEDGEMENTS

This work was supported by the Division of Intramural Research, NIAID. Glennys Reynoso provided outstanding technical support. Thanks to Jia Jie Wei, Mina Seedhom, and Davide Angeletti for productive conversations. Thanks also to Zhiping Ye at FDA and the Biodefense and Emerging Infections Research Resources Repository (BEI Resources) for viruses.

REFERENCES

Alymova, I.V., York, I.A., Air, G.M., Cipollo, J.F., Gulati, S., Baranovich, T., Kumar, A., Zeng, H., Gansebom, S., McCullers, J.A., 2016. Glycosylation changes in the globular head of H3N2 influenza hemagglutinin modulate receptor binding without affecting virus virulence. Sci Rep 6, 36216.

An, Y., McCullers, J.A., Alymova, I., Parsons, L.M., Cipollo, J.F., 2015. Glycosylation Analysis of Engineered H3N2 Influenza A Virus Hemagglutinins with Sequentially Added Historically Relevant Glycosylation Sites. Journal of proteome research 14, 3957-3969.

Aytay, S., Schulze, I.T., 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.

Blackburne, B.P., Hay, A.J., Goldstein, R.A., 2008. Changing selective pressure during antigenic changes in human influenza H3. PLoS Pathog 4, e1000058.

CDC, 2017. Vaccine Effectiveness - How Well Does the Flu Vaccine Work?

Couch, R.B., Kasel, J.A., 1983. Immunity to influenza in man. Annu Rev Microbiol 37, 529-549.

Daniels, R., Kurowski, B., Johnson, A.E., Hebert, D.N., 2003. N-linked glycans direct the cotranslational folding pathway of influenza hemagglutinin. Mol Cell 11, 79-90.

Das, S.R., Hensley, S.E., David, A., Schmidt, L., Gibbs, J.S., Puigbo, P., Ince, W.L., Bennink, J.R., Yewdell, J.W., 2011. Fitness costs limit influenza A virus hemagglutinin glycosylation as an immune evasion strategy. Proc Natl Acad Sci U S A 108, E1417-1422.

de Graaf, M., Fouchier, R.A., 2014. Role of receptor binding specificity in influenza A virus transmission and pathogenesis. EMBO J 33, 823-841.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460-2461.

Eggink, D., Goff, P.H., Palese, P., 2014. Guiding the Immune Response against Influenza Virus Hemagglutinin toward the Conserved Stalk Domain by Hyperglycosylation of the Globular Head Domain. Journal of Virology 88, 699-704.

Gupta, R.B.S., 2004. Prediction of N-glycosylation sites in human proteins.

Hrincius, E.R., Liedmann, S., Finkelstein, D., Vogel, P., Gansebom, S., Samarasinghe, A.E., You, D., Cormier, S.A., McCullers, J.A., 2015. Acute Lung Injury Results from Innate Sensing of Viruses by an ER Stress Pathway. Cell Rep 11, 1591-1603.

Job, E.R., Deng, Y.M., Barfod, K.K., Tate, M.D., Caldwell, N., Reddiex, S., Maurer-Stroh, S., Brooks, A.G., Reading, P.C., 2013. Addition of glycosylation to influenza A virus hemagglutinin modulates antibody-mediated recognition of H1N1 2009 pandemic viruses. J Immunol 190, 2169-2177.

Keil, W., Geyer, R., Dabrowski, J., Dabrowski, U., Niemann, H., Stirm, S., Klenk, H.D., 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.

Khatri, K., Klein, J.A., White, M.R., Grant, O.C., Leymarie, N., Woods, R.J., Hartshorn, K.L., 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.

Medina, R.A., Stertz, S., Manicassamy, B., Zimmermann, P., Sun, X., Albrecht, R.A., Uusi-Kerttula, H., Zagordi, O., Belshe, R.B., Frey, S.E., Tumpey, T.M., Garcia-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, 187ra170.

Park, S., Lee, I., Kim, J.I., Bae, J.Y., Yoo, K., Kim, J., Nam, M., Park, M., Yun, S.H., Cho, W.I., Kim, Y.S., Ko, Y.Y., Park, M.S., 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.

Rambaut, A., Drummond, A., 2009. FigTree v1. 3.1.

Rozo, M., Gronvall, G.K., 2015. Reply to "The 1977 H1N1 Influenza Virus Reemergence Demonstrated Gain-of-Function Hazards". MBio 6, e01524-01515.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7, 539.

Sun, H., Yang, J., Zhang, T., Long, L.-P., Jia, K., Yang, G., Webby, R.J., Wan, X.-F., 2013a. Using Sequence Data To Infer the Antigenicity of Influenza Virus. mBio 4, e00230-00213.

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.

Sun, X., Jayaraman, A., Maniprasad, P., Raman, R., Houser, K.V., Pappas, C., Zeng, H., Sasisekharan, R., Katz, J.M., Tumpey, T.M., 2013b. 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.

Tate, M.D., Job, E.R., Deng, Y.M., Gunalan, V., Maurer-Stroh, S., Reading, P.C., 2014. Playing hide and seek: how glycosylation of the influenza virus hemagglutinin can modulate the immune response to infection. Viruses 6, 1294-1316.

Vigerust, D.J., Ulett, K.B., Boyd, K.L., Madsen, J., Hawgood, S., McCullers, J.A., 2007. N-linked glycosylation attenuates H3N2 influenza viruses. J Virol 81, 8593-8600.

Vijaykrishna, D., Holmes, E.C., Joseph, U., Fourment, M., Su, Y.C., Halpin, R., Lee, R.T., Deng, Y.M., Gunalan, V., Lin, X., Stockwell, T.B., Fedorova, N.B., Zhou, B., Spirason, N., Kuhnert, D., Boskova, V., Stadler, T., Costa, A.M., Dwyer, D.E., Huang, Q.S., Jennings, L.C., Rawlinson, W., Sullivan, S.G., Hurt, A.C., Maurer-Stroh, S., Wentworth, D.E., Smith, G.J., Barr, I.G., 2015. The contrasting phylodynamics of human influenza B viruses. Elife 4, e05055.

WHO, 2003. Influenza Fact Sheet No. 211.

Yang, H., Carney, P.J., Chang, J.C., Guo, Z., Villanueva, J.M., Stevens, J., 2015. Structure and receptor binding preferences of recombinant human A(H3N2) virus hemagglutinins. Virology 477, 18-31.

Zhang, Y., Aevermann, B.D., Anderson, T.K., Burke, D.F., Dauphin, G., Gu, Z., He, S., Kumar, S., Larsen, C.N., Lee, A.J., Li, X., Macken, C., Mahaffey, C., Pickett, B.E., Reardon, B., Smith, T., Stewart, L., Suloway, C., Sun, G., Tong, L., Vincent, A.L., Walters, B., Zaremba, S., Zhao, H., Zhou, L., Zmasek, C., Klem, E.B., Scheuermann, R.H., 2017. Influenza Research Database: An integrated bioinformatics resource for influenza virus research. Nucleic Acids Res 45, D466-D474.

Hemagglutinin Glycan Clock Guides Human Influenza A Virus Evolution

SUPPLEMENTAL MATERIAL

Material and Methods

fig. S1-S6

Table S1

MATERIALS AND METHODS

Virus preparation and purification

We propagated each virus by infecting 5, 10-day old fertilized chicken eggs with 100µL of viral stock diluted 1:2000 in PBS. Infected eggs were incubated at 34.5˚C and 65% humidity for ~41h. We first clarified the allantoic fluid (AF) by centrifuging 4500 x g for 20min, then pelleted the virus by centrifuging for 2h at 26,000 x g. After incubating the viral pellet in 1.7mL PBS overnight at 4˚C, we layered the virus onto a discontinuous 15%-60% sucrose gradient and spun at 34,000 x g for 2h. We collected virus at the gradient interface and spun it down a final time in PBS at 34,000 x g for 2h. Virus was resuspended in 100µL PBS and stored as milky opalescent suspensions long term at 4˚C. Hemagglutination titers for each virus were measured from neat AF. A complete list of viruses and their origin is found in **File S1**. Total protein content of each virus prep was determined by coomassie-based assay performed per manufacturer instructions [Bio-Rad, DC Assay].

Allantoic fluid (AF) from eggs infected with 93 out of the 97 viruses we attempted to grow showed hemagglutination activity (HAU), indicating viral presence. We collected pure virus from all viruses in amounts ranging from 20-6920 µg. Sham infection and purification of AF from control eggs did not produce enough protein for detection, and ran cleanly on a gel (10 µL of a 100 µL prep, Lane #29, **fig S1**).

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 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 Igs to identify which bands corresponded to which proteins on the coomassie gels (**fig S2**).

Protein gels and immunoblotting

We mixed purified virus (1 µg protein) with 4× NuPage loading buffer (Invitrogen), and boiled for 10 min at 96˚C. We electrophoresed samples with Chameleon Duo Li-Cor ladder on 4–12% 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, shaking at RT. After removing fixative, we added 10 ml GelCode stain (Pierce) and shook for 30 min at RT, then destained the gels with water overnight. For immunoblotting, we transferred proteins from gels to PVDF membranes with the iBLOT at P3 setting for 7 min. We blocked membranes for 30 min at room temperature (RT) StartingBlock (Thermo). After incubating with primary Ig or sera for 1 hr at RT, and washing 5x for 5 min each in TBST (10 mM Tris, 150 mM NaCl, 0.1% Tween-20), we added secondary Ig, repeating the washing step incubation. We imaged blots and commassie gels on a Li-Cor Odyssey. Band signals and molecular weights were calculated using Li-Cor's ImageStudio software.

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 varied from the mean molecular weight of M1 and NP proteins on each gel (**fig. S3**). Linear regression, Pearson coefficient, and P-values were calculated using PRISM software.

Influenza sequences and N-glycosylation prediction

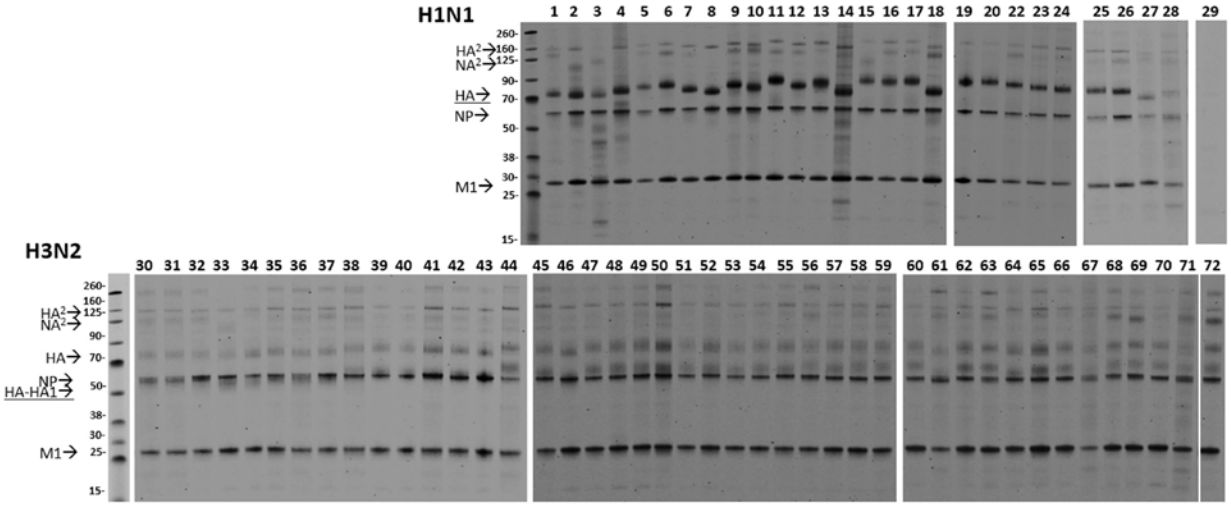
Human HA protein sequences were retrieved from the NIAID Influenza Research Database (IRD) (Zhang et al., 2017) through the web site at http://www.fludb.org (accessed April 20, 2017). Sequences with identical strain names were retained only if the majority of members had identical sequences. Sequences were aligned to A/California/04/2009(H1N1) (FJ966082) with MAFFT v7.305b, and 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 with both H1 and H3 numbering conventions is shown in **Table S1**. Potential glycosylation sites were identified by searching for the NX[S/T]Y motif where X/Y were any amino acid other than proline or by using NetNGlyc 1.0 artificial neural network-based prediction software (Gupta, 2004). For the quantitative analysis in **Fig. 1** and **fig. S4**, 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 with “N--" were counted as negative.

pH1N1 Sweeper Strains bioinformatic analysis

All pH1N1 in FluDB from 2009 to present day were collected and annotated with collection country, date, and glycosylation status. Sequences glycosylated at N179 from 2009-2015 were aligned by ClustalΩ (Sievers et al., 2011). An unrooted phylogenetic tree of all 365, N179-glycosylated sequences was made using FigTree (Rambaut and Drummond, 2009). The earliest strains in the same clade as the SWSs were identified as: A/Shiraz/4/2015, A/Shiraz/3/2015, A/Taipei/0021/2015, A/Adana/08/2015, A/Shiraz/6/2015.

To identify any additional mutations occurring with N179 we identified all sequences with >99% sequence similarity to the 6 earliest SWSs using USEARCH (max ~17nt differences) (Edgar, 2010). This represented 15% of all FluDB sequences. An unrooted phylogenetic tree of a downsampled set of 370 sequences representing pre-SWS and SWSs (as well as outgroup A/California/07/2009) was aligned in ClustalΩ is shown to the right. From this alignment, the two conserved amino acid changes that define SWSs, S179N and I233T were clearly seen.

Fig. S1.



Lane	Year	H1N1
1	1931	A/Swine/1931 (H1N1)
2	1933	A/Wilson-Smith/1933 (H1N1) (clone)
3	1934	A/Puerto Rico/8/1934 (H1N1) Mca Clone
4	1935	A/Melbourne/1935 (H1N1)
5	1940	A/Hickox/1940 (H1N1)
6	1942	A/Bellamy/1942 (H1N1)
7	1943	A/Weiss/1943 (H1N1)
8	1946	A/Melbourne/1/1946 (H1N1) (clone)
9	1946	A/Cameron/1946 (H1N1)
10	1947	A/Fort Monmouth/1/1947 (H1N1) (clone)
11	1951	A/Fort Leonard Wood/1951 (H1N1)
12	1954	A/Malaysia/302/1954 (H1N1)
13	1957	A/Denver/1/1957 (H1N1)
14	1976	A/New Jersey/8/1976 (swine origin H1N1)
15	1977	A/USSR/90/1977 (HA) xPR8 (NA) (H1N1)
16	1978	A/Brazil/11/1978 (HA+NA) xPR8 = H1N1
17	1980	A/India/6263/1980 (HA+NA) xPR8 = H1N1
18	1982	A/Memphis/4/1982 (swine origin H1N1)
19	1983	A/Chile/1/1983 (HA+NA) xPR8 = H1N1
20	1983	A/Dunedin/6/1983 (HA+NA) xPR8 = H1N1
22	1991	A/Texas/36/1991 (H1N1)
23	1995	A/Beijing/262/1995 (H1N1)
24	1999	A/New_Caledonia/20/1999 (H1N1)
25	2006	A/Solomon Islands/3/2006 (H1N1)
26	2007	A/Brisbane/59/2007 (H1N1)
27	2009	A/California/07/2009 (pH1N1)
28	2012	A/Georgia/M5081/2012 (pH1N1)
29		Allantoic Fluid Control

Lane	Year	H3N2
30	1968	A/Aichi/2/1968 (H3N2)
31	1968	A/Hong Kong/1/1968 (H3N2) (mother clone)
32	1968	A/Northern_Territory/60/1968 (H3N2)
33	1971	A/Hong Kong/107/1971 (HA+NA) xPR8 = H3N2
34	1972	A/Hong Kong/5/1972 (HA+NA) xPR8 = H3N2
35	1972	A/England/42/1972 (HA+NA) xPR8 = H3N2
36	1972	A/Udorn/307/1972 (H3N2)
37	1974	A/Georgia/101/1974 (HA+NA) xPR8 = H3N2
38	1974	A/Scotland/840/1974 (HA+NA) xPR8 = H3N2
39	1975	A/England/864/1975 (HA+NA) xPR8 = H3N2
40	1975	A/Tokyo/1/1975 (HA+NA) xPR8 = H3N2
41	1975	A/Victoria/3/1975 (HA+NA) xPR8 = H3N2
42	1976	A/Victoria/112/1976 (HA+NA) xPR8 = H3N2
43	1979	A/Bangkok/1/1979 (HA+NA) xPR8 = H3N2
44	1982	A/Philippines/2/1982 (H3N2) (clone)
45	1984	A/Caen/1/1984 (HA+NA) xPR8 = H3N2
46	1985	A/Mississippi/1/1985 (HA+NA) xPR8 = H3N2
47	1986	A/Leningrad/360/1986 (HA+NA) xPR8 = H3N2
48	1987	A/Sichuan/2/1987 (HA+NA) xPR8 = H3N2
49	1987	A/Shanghai/11/1987 (HA+NA) xPR8 = H3N2 X-99
50	1988	A/England/427/1988 (HA+NA) xPR8 = H3N2
51	1989	A/Sichuan/60/1989 (HA+NA) xPR8 = H3N2 Mutant NA+
52	1989	A/Beijing/4/1989 (HA+NA) xPR8 = H3N2
53	1989	A/Guizhou/54/1989 (HA+NA) xPR8 = H3N2
54	1989	A/Guangdong/39/1989 (HA+NA) xPR8 = H3N2
55	1990	A/Shanghai/24/1990 (HA+NA) xPR8 = H3N2 High Yield
56	1992	A/Harbin/15/1992 (HA+NA) xPR8 = H3N2
57	1992	A/Beijing/32/1992 (HA+NA) xPR8 = H3N2
58	1993	A/Ann Arbor/3/1993 (HA+NA) xPR8 = H3N2
59	1993	A/Shangdong/9/1993 (HA+NA) xPR8 = H3N2
60	1994	A/Johannesburg/33/1994 (HA+NA) xPR8 = H3N2
61	1995	A/Nanchang/933/1995 (H3N2)
62	1996	A/South Africa/1147/1996 (HA+NA) xPR8 = H3N2
63	1997	A/Sydney/5/1997 (H3N2)
64	1999	A/Moscow/10/1999 (HA+NA) xPR8 = H3N2
65	1999	A/California/32/1999 (HA+NA) xPR8 = H3N2
66	2000	A/Ulan Ude/01/2000 (HA+NA) xPR8 = H3N2
67	2005	A/Wisconsin/67/2005 (H3N2)
68	2007	A/Brisbane/10/2007 (H3N2)
69	2007	A/Uruguay/716/2007 (H3N2)
70	2009	A/Perth/16/2009 (HA+NA) xPR8 = H3N2 NIB-64
71	2009	A/Wisconsin/15/2009 (H3N2)
72	2011	A/Victoria/361/2011 (H3N2)

Fig. S1. SDS-PAGE of purified virus. Bands used for size quantitation are labeled. Representative blots are shown. Lane identifications with reassortant status are shown below blots.

Fig. S2.

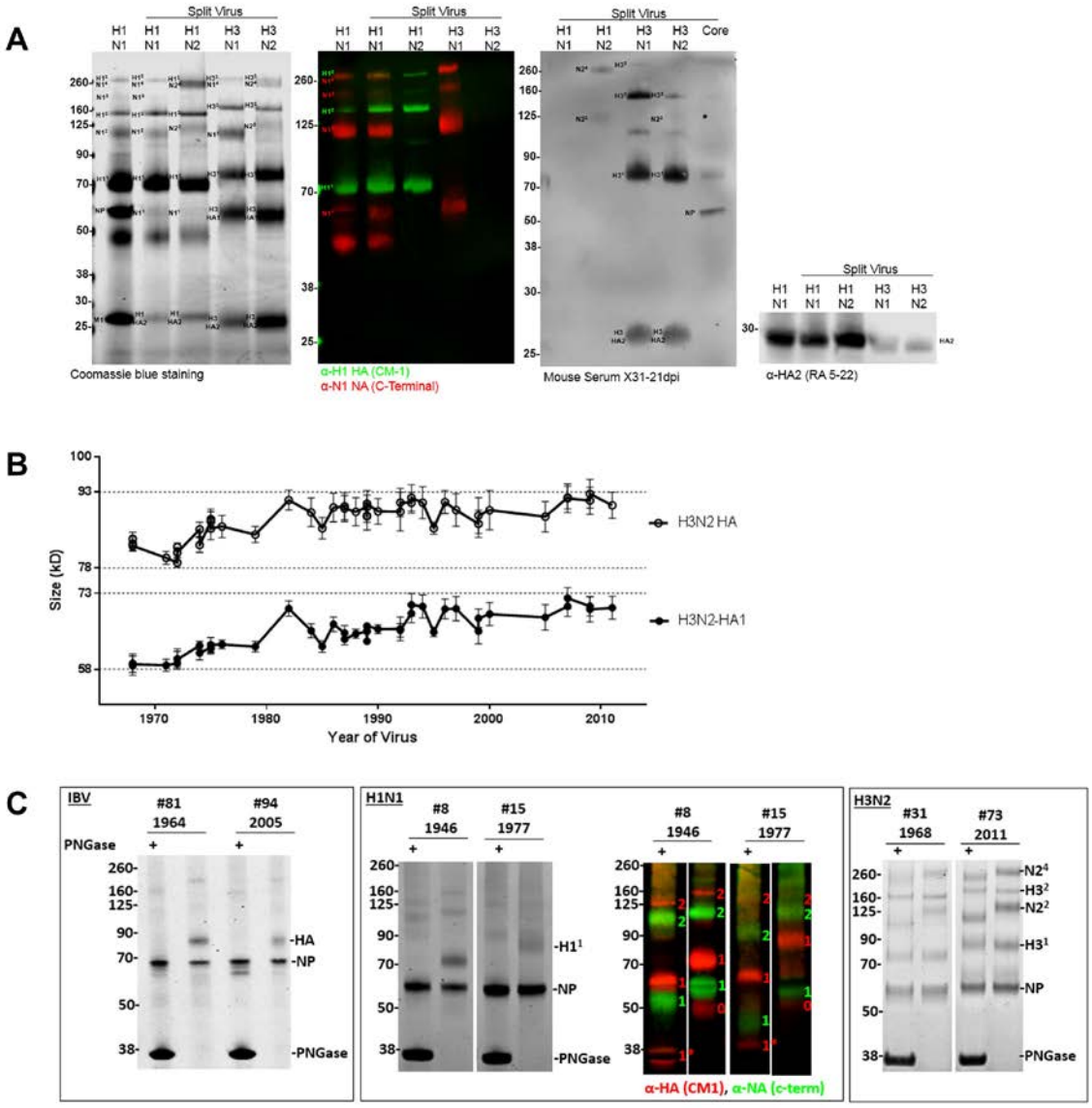


Fig. S2. Identification of viral proteins in Coomassie blue stained gels. **(A)** Purified whole PR8 or reassortant, detergent split viruses were stained with Coomassie blue; probed with monoclonal antibodies against HA1 (CM1), N1 (C-terminal), and HA2 (RA-5 22); or convalescent sera from a H3N2 reassortant infected mouse. **(B)** H3N2 HA monomer and H3N2-HA1 band sizes correlate almost perfectly. As the resolution was better for H3N2-HA1, these were used for analysis. **(C)** PNGase treatment of viruses shows deglycosylation of glycoproteins for representative viruses. Coomassie stained gels and monoclonal antibodies against H1 (CM1) and N1 (C-terminal) for H1N1.

Fig. S3.

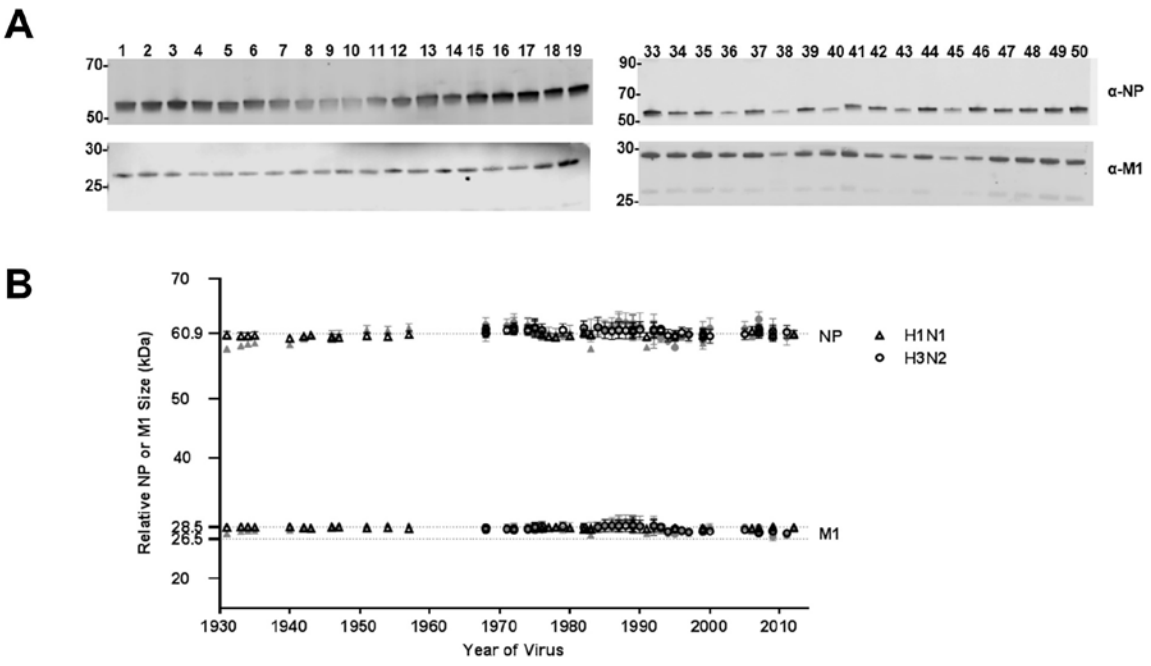


Fig. S3. NP and M1 proteins serve as internal ruler. (A) Western blots of NP and M1 for representative viruses. **(B)** Quantitation of NP and M1 relative size from SDS-PAGE. Raw data (grey) was normalized to account for gel distortion (open black symbols). Average size of each protein is shown on the Y-axis. Error bars are standard error from n = 3-6 blots.

Fig. S4.

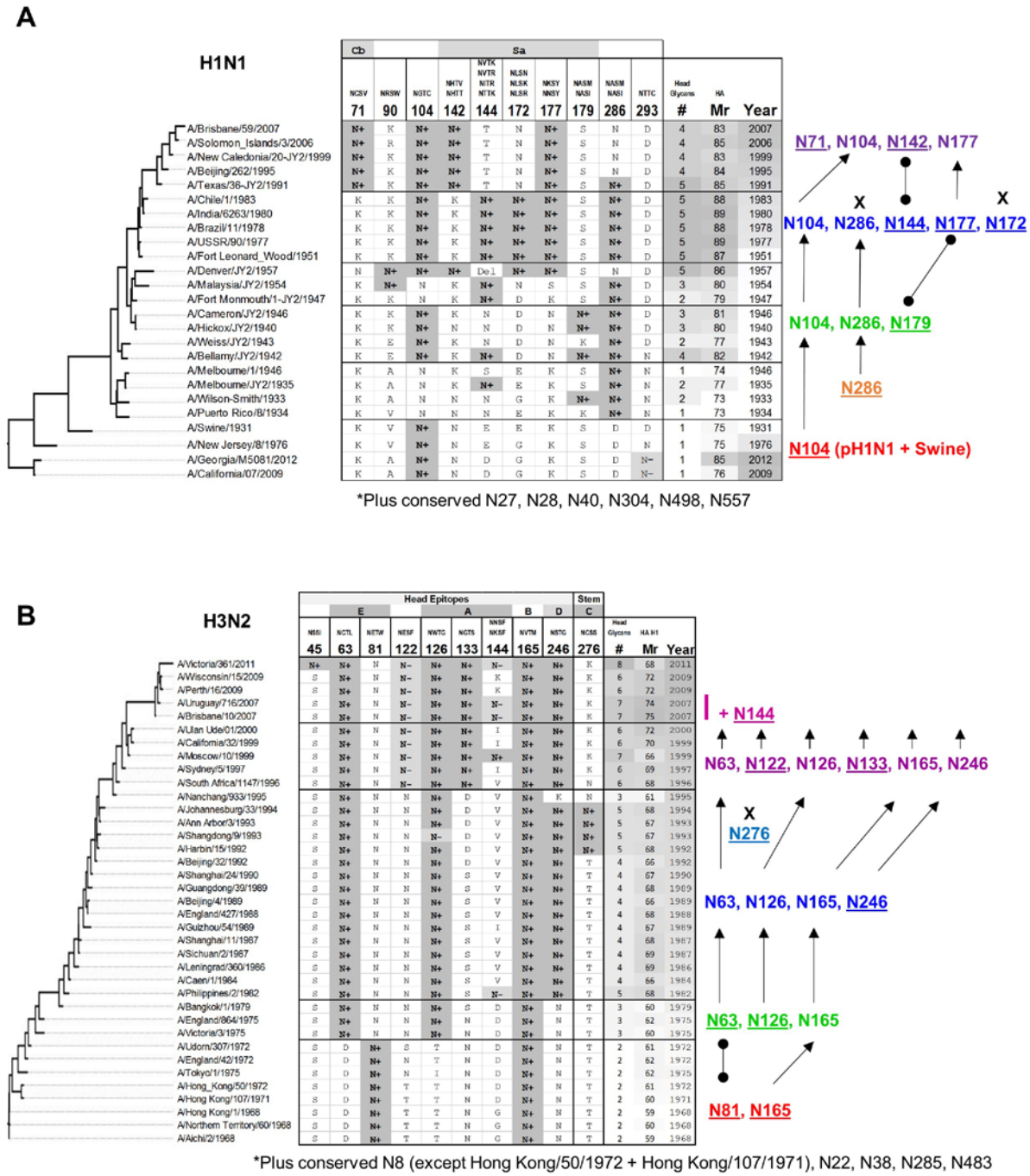


Fig. S4. Glycan groups of IAV viral panel. (A) Phylogeny of H1N1 or **(B)** H3N2 viruses and NetNGlyc predicted glycosylation status of HA Asn (N+ or N-), the substituted residue present, or a deletion (Del). The total number of head glycans, average HA size, and year of virus collection is shown for each virus. The glycosylation progression is shown color coded to match individual viruses in **Fig. 1B**. New glycosylation sites are underlined. Arrows connect conserved glycosylation sites and barbells connect glycans in the same epitope, but not same site. Lost glycosylation sites are noted with an X. Conserved glycans among all viruses are listed below the table.

Fig. S5.

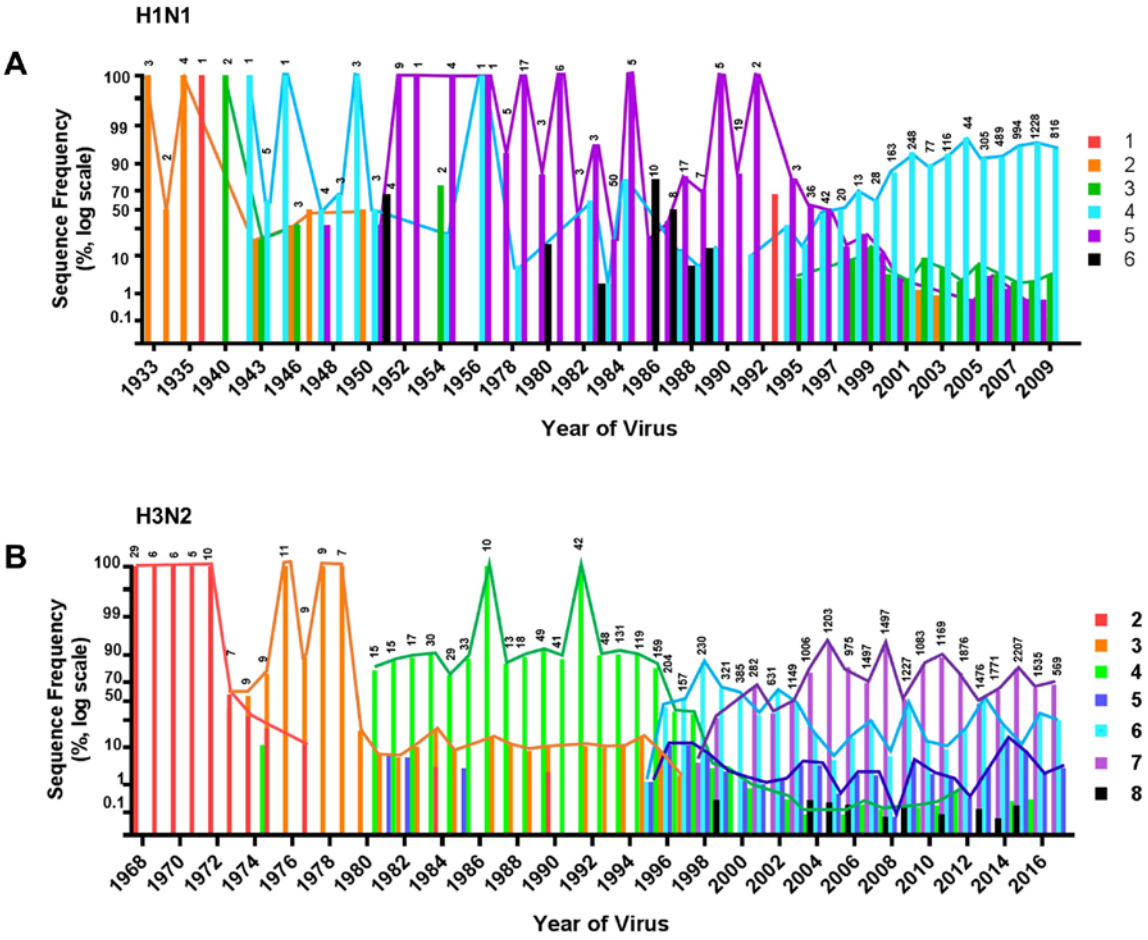


Fig. S5. Frequency of total sequences by number of glycosylation sites. (A) Percentage of H1N1 or **(B)** H3N2 sequences in FluDB with indicated number of predicted head glycans. Above each lane are total number of sequences in FluDB from that year. Y-Axis is log scale. Sequences above the glycan limit 6 for H1N1 and 8 for H3N2 are shown in black.

Fig. S6.

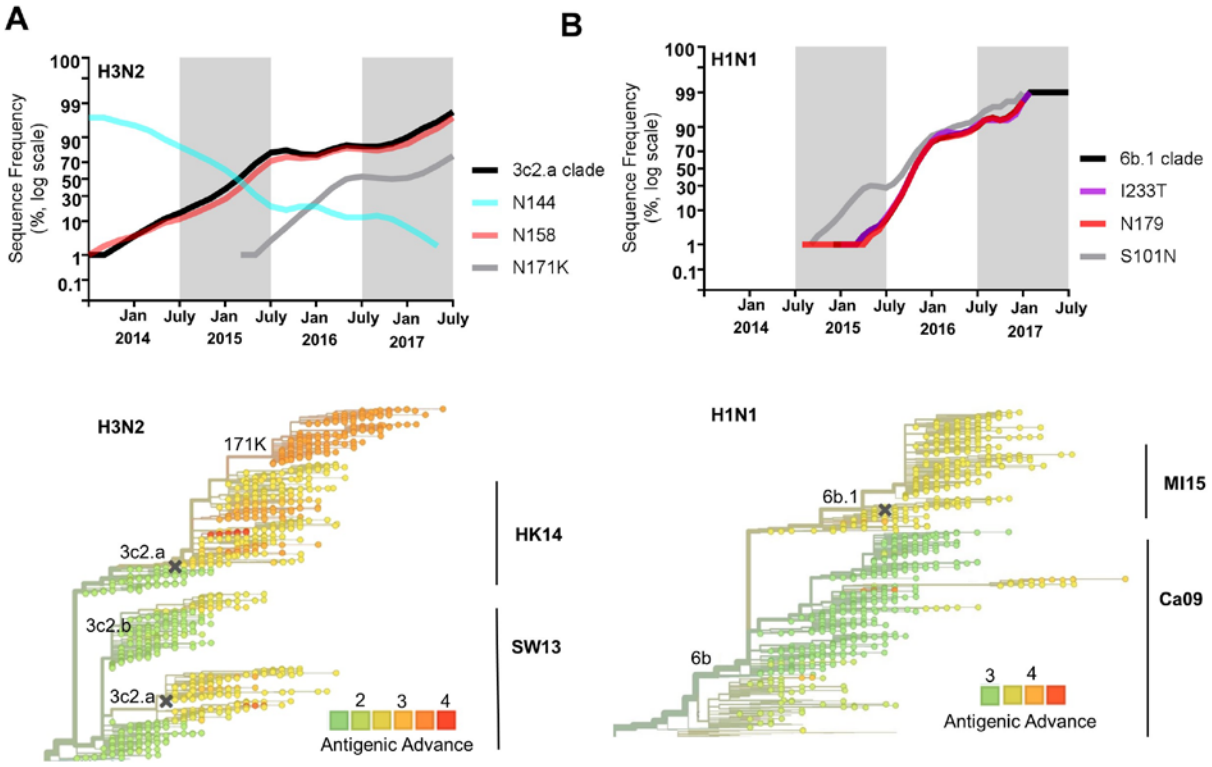


Fig. S6. Major IAV clades track with glycan evolution. Using nextflu.org we tracked the most recent glycan changes for **(A)** H1N1 and **(B)** H3N2 viruses in terms of frequency of the major clade (black) and other mutations of interest (top). On bottom, phylogenic trees are colored according to antigenic advance measured as $\log_2$ HI titer distance from the root of the tree. Vaccine strains are marked with an X. Antigenic clusters are labelled to the right.

Table S1. Relevant amino acid residues with H1 and H3 numbering. Calculated using the "HA subtype numbering conversion tool (beta)" on FluDB. Numbering in our text starts from the initial HA methionine (M=1).

In Paper	M=1	H1	H3
N27	27	10	20
N28	28	11	21
N40	40	23	33
N71	71	54	63
N90	90	73	82
S101N	101	84	92
N104	104	87	94
N136	136	119	125.1
N142	142	125	129
N144	144	127	131
N172	172	155	158
N177	177	160	163
N179	179	162	165
I233T	233	216	219
N286	286	269	271
N304	304	287	289
N498	498	481	483
N557	557	HA2 213	541