Meta Gene 27 (2021) 100828

Contents lists available at ScienceDirect

Meta Gene

journal homepage: www.elsevier.com/locate/mgene

Effect of *N*-linked glycosylation at position 162 of hemagglutinin in influenza A virus A(H1N1)pdm09

Yoshiyuki Suzuki [a,*], Maho Nishimura [a], Tatsuya Inoue [a], Yuki Kobayashi [b]

[a] Graduate School of Science, Nagoya City University, Nagoya-shi, Aichi-ken 467-8501, Japan
[b] Nihon University Veterinary Research Center, Fujisawa-shi, Kanagawa-ken 252-0880, Japan

ARTICLE INFO

Keywords:
Hemagglutinin
Influenza A virus
N-linked glycosylation
Nonsynonymous diversity
Synonymous diversity

ABSTRACT

Subtype H1N1 influenza A virus (IAV) that caused a pandemic in 2009, called A(H1N1)pdm09, has been circulating as a seasonal IAV among humans, escaping from herd immunity through amino acid substitutions in hemagglutinin (HA). Additionally, position 162 of HA in A(H1N1)pdm09 evolved to be *N*-linked glycosylated by 2017. Here the effect of *N*-linked glycosylation at position 162 of HA in A(H1N1)pdm09 was examined by comparing the ratio of nonsynonymous diversity (π_N) to synonymous diversity (π_S) (π_N/π_S) at structurally defined categories of amino acid sites between 1785 (162+) and 7562 (162−) sequences of HA that were and were not considered to be *N*-linked glycosylated at position 162, respectively. In 162− sequences, π_N/π_S for the sites located inside 15 Å from position 162 was not significantly smaller than one and was greater than that for the sites located outside 15 Å. In 162+ sequences, π_N/π_S for the sites located inside and outside 15 Å from position 162 were smaller than the corresponding values in 162− sequences, and were not significantly different from each other. These results suggest that *N*-linked glycosylation at position 162 of HA may shield epitopes against antibodies in A(H1N1)pdm09.

Influenza A virus (IAV) is an etiological agent of influenza. The virion of IAV is enveloped and 80–120 nm in diameter. The genome of IAV consists of eight segments of linear and single-stranded RNA with negative polarity. Hemagglutinin (HA) and neuraminidase (NA) constitute envelope glycoproteins, with HA existing ~10 times more abundantly than NA (Mitnaul et al., 1996). HA is the sialic acid receptor-binding protein and the major target of humoral immunity (Skehel and Wiley, 2000). NA is the sialidase for cleaving the sialic acid receptor from galactose (McAuley et al., 2019). The balance between the activities of HA and NA (HA-NA balance) is crucial for efficient replication of IAV (Wagner et al., 2002; Xu et al., 2012). According to antigenic and genetic properties of HA and NA, IAV is classified into subtypes H1-H18 and N1-N11, respectively.

Subtype H3N2 IAV that caused a pandemic in 1968, called A(H3N2)pdm68 in the present study, subsequently became seasonal and has been circulating among humans. A(H3N2)pdm68 has been escaping from herd immunity through amino acid substitutions in HA (Suzuki, 2006; Chen and Sun, 2011; Murrell et al., 2012) and reassortments involving the HA segment (Barr et al., 2005; Nelson and Holmes, 2007; Suzuki, 2010). Additionally, the number of sequons, which are defined as the sequence of asparagine (N), any amino acid (X) except for proline (P), and serine (S) or threonine (T) increased in HA during evolution of A(H3N2)pdm68. The N residues of sequons in HA were experimentally demonstrated to be mostly *N*-linked glycosylated (Altman et al., 2019). It is considered that *N*-linked glycans on HA provided A(H3N2)pdm68 with selective advantage through shielding epitopes against antibodies (Suzuki, 2011; Kobayashi and Suzuki, 2012b). Notably, *N*-linked glycans may also interfere with the sialic acid receptor-binding activity of HA (Wang et al., 2009; de Vries et al., 2010). However, in parallel with the increase in the number of sequons, the net-charge, which is calculated by subtracting the number of aspartic acid (D) and glutamic acid (E) residues from the number of histidine (H), lysine (K), and arginine (R) residues and is usually positive for HA, also increased during evolution of A(H3N2)pdm68 (Hensley et al., 2009). Since the sialic acid receptor is negatively charged, the increase in the net-charge possibly compensated for the reduction in the sialic acid receptor-binding activity of HA caused by the increase in the number of sequons (Kobayashi and Suzuki, 2012a).

Besides, subtype H1N1 IAV that caused a pandemic in 2009, called A(H1N1)pdm09, subsequently became seasonal and has been circulating among humans. Similarly to the case for A(H3N2)pdm68, A(H1N1)pdm09 has been escaping from herd immunity through amino acid

* Corresponding author at: Graduate School of Science, Nagoya City University, 1 Yamanohata, Mizuho-ku, Nagoya-shi, Aichi-ken 467-8501, Japan.
E-mail address: yossuzuk@nsc.nagoya-cu.ac.jp (Y. Suzuki).

https://doi.org/10.1016/j.mgene.2020.100828
Received 20 August 2020; Received in revised form 13 September 2020; Accepted 6 November 2020
Available online 11 November 2020

substitutions in HA (Koel et al., 2015; Lee et al., 2015), although the evolutionary significance of reassortment has not been clarified (Su et al., 2015). Interestingly, the number of sequons also increased in HA during evolution of A(H1N1)pdm09; a novel sequon was created at positions 162–164 in 2014 and became prevalent by 2017 (Altman et al., 2019). The purpose of the present study was to examine the effect of *N*-linked glycosylation at position 162 of HA in A(H1N1)pdm09 from molecular evolutionary analyses of nucleotide sequences.

HA is a homotrimeric type 1 transmembrane glycoprotein. HA of A(H1N1)pdm09 comprises 566 amino acid sites, constituting the signal peptide (17 sites: positions [−17]–[−1]), HA1 (327 sites: positions 1–327), and HA2 (222 sites: positions 328–549) from the N-terminus to the C-terminus (Skehel and Wiley, 2000). The ectodomain of HA trimer consists of a distal globular head and a proximal stem. The globular head is composed of a part of HA1 (positions 43–274), while the stem is composed of the remaining parts of HA1 (positions 1–42 and 275–327) and a part of HA2 (positions 328–503) (Lu et al., 2014). The sialic acid receptor-binding sites reside in the globular head, and are surrounded by neutralization epitopes in the globular head as well as the stem of HA1 (Yan et al., 2019).

Nucleotide sequences encoding the entire coding regions of HA (566 amino acid sites) and NA (469 amino acid sites) without ambiguous nucleotides, minor gaps, or premature termination codons for 9637 strains of A(H1N1)pdm09 were retrieved from International Nucleotide Sequence Database (INSD) through NIAID Influenza Research Database (IRD) (Zhang et al., 2017) on March 7, 2020 (Supplementary Table S1). Isolation years of these strains ranged from 2009 to 2019; the numbers of strains isolated in 2009–2019 were 4294, 1114, 742, 307, 633, 447, 411, 1301, 349, 36, and 3, respectively. The number of sequons and the net-charge were calculated for the deduced amino acid sequences of HA and NA, and were averaged for each isolation year.

In HA, the N residues of the sequons identified in any of 9637 sequences were located at positions 10, 11, 23, 38, 87, 119, 147, 160, 162, 183, 205, 239, 276, 287, 308, 431, 438, 472, 481, and 540 (Supplementary Table S2). To focus on the effect of *N*-linked glycosylation at position 162, inconsistencies in the configuration of other sequons were removed by eliminating the sequences that did not contain the sequons with the N residues located at positions 10, 11, 23, 87, 276, 287, 481, and 540 and those that contained the sequons with the N residues located at positions 38, 119, 147, 160, 183, 205, 239, 308, 431, 438, and 472.

It was experimentally demonstrated that position 162 of the sequon at positions 162–164 of HA in A(H1N1)pdm09 was *N*-linked glycosylated only when positions 163 and 216 were occupied by glutamine (Q) and threonine (T), respectively (Altman et al., 2019). Therefore, the HA sequences that contained a sequon at positions 162–164 with the Q and T residues at positions 163 and 216, respectively, were regarded to be *N*-linked glycosylated at position 162 and thus were designated 162+ sequences, and the HA sequences that did not contain a sequon at positions 162–164 were regarded not to be *N*-linked glycosylated at position 162 and thus were designated 162− sequences in the present study. The HA sequences that contained a sequon at positions 162–164 without the Q and T residues at positions 163 and 216, respectively, were eliminated from the analysis for clarity.

Effect of *N*-linked glycosylation at position 162 on other amino acid sites of HA may be evaluated by comparing the ratio of nonsynonymous diversity (π_N) to synonymous diversity (π_S) (π_N/π_S) between 162− and 162+ sequences. For this purpose, it is required that 162− and 162+ sequences form distinct clusters in the phylogenetic tree. Therefore, 162− and 162+ sequences were aligned together with the HA sequence for A/South Carolina/1/1918(H1N1) (INSD accession number: AF117241), which was adopted as the outgroup, using the computer program MAFFT (version 7.305b) (Katoh et al., 2002). The HA tree was constructed by the p distance-based neighbor-joining (NJp) method, which was reported to perform better than other methods generally in constructing the phylogenetic tree (Nei and Kumar, 2000; Yoshida and Nei, 2016), using MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms (version 10.1.7) (Kumar et al., 2018). The reliabilities of interior branches in the HA tree were assessed by the bootstrap method with 1000 resamplings. After eliminating 162− sequences included in the cluster of 162+ sequences and *vice versa*, 7562 162− and 1785 162+ sequences (totally 9347 sequences) of HA for A(H1N1)pdm09 were observed to form distinct clusters in the HA tree, although the bootstrap probability for the interior branch supporting the cluster of 162+ sequences was not significantly high (66%) (Supplementary Fig. S1) (Rambaut, 2018).

Comparison of π_N/π_S between 162− and 162+ sequences was conducted for structurally defined categories of amino acid sites based on the tertiary structure of HA trimer for A(H1N1)pdm09 (A/Washington/5/2011(H1N1); Protein Data Bank (PDB) accession number: 4LXV). The amino acid sites in HA1 and HA2 were categorized into solvent accessible and inaccessible sites using PISA (version 1.48) (Krissinel and Henrick, 2007). Reportedly, the sites located within 8–17 Å from the N residue of sequon were considered to be covered by *N*-linked glycan (Kobayashi and Suzuki, 2012b). Therefore, the sites whose α-carbons were located within 10 Å or 15 Å from the α-carbon of the amino acid at position 162 in HA trimer were assumed to be covered by *N*-linked glycan in 162+ sequences (Supplementary Table S3). For each category of sites, π_N/π_S was computed for 162− and 162+ sequences using MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms (version 10.1.7) (Kumar et al., 2018). The variances of π_N and π_S were obtained by the bootstrap method with 1000 resamplings, and the variance of π_N/π_S was obtained by the formula, $V(\pi_N/\pi_S) = (1/\pi_S^2)V(\pi_N) + (\pi_N^2/\pi_S^4)V(\pi_S)$ (Ota and Nei, 1994).

Occurrences of reassortments during evolution of A(H1N1)pdm09 may be inferred from differences in topologies of phylogenetic trees between different genomic segments (Suzuki, 2010). Therefore, NA sequences for the same set of 9347 strains of A(H1N1)pdm09 employed for constructing the HA tree above were aligned together with the NA sequence for A/South Carolina/1/1918(H1N1) (INSD accession number: AF250356), which was adopted as the outgroup, using MAFFT (version 7.305b) (Katoh et al., 2002). The NA tree was constructed by the NJp method, and the reliabilities of interior branches were assessed by the bootstrap method with 1000 resamplings using MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms (version 10.1.7) (Kumar et al., 2018). The topology of the NA tree was compared with that of the HA tree (Supplementary Fig. S1).

In the deduced amino acid sequences of HA for A(H1N1)pdm09 isolated during 2009–2019, both the number of sequons and the net-charge increased with the isolation year; the correlation coefficient (r) for the number of sequons and the isolation year was 0.888 ($P < .05$ by t-test) and the r for the net-charge and the isolation year was 0.784 ($P < .05$ by t-test) (Supplementary Fig. S2). The number of sequons and the net-charge were positively correlated with each other ($r = 0.544$; $P < .05$ by t-test). In the deduced amino acid sequences of NA, the net-charge increased with the isolation year ($r = 0.599$; $P < .05$ by t-test) (Supplementary Fig. S2). However, the number of sequons did not increase with the isolation year ($r = -0.306$) (Sun et al., 2011; Zhang et al., 2004), and thus the number of sequons and the net-charge were not positively correlated with each other ($r = 0.000463$).

Estimate of π_N/π_S for the entire region of HA was smaller than one in both 162− and 162+ sequences of A(H1N1)pdm09 ($P < .05$ by Z-test), reflecting functional constraint operating on HA (Fig. 1; Supplementary Table S4). However, when HA was divided into HA1 and HA2, π_N/π_S for HA1 was greater than that for HA2 in both 162− and 162+ sequences, apparently because mutations in HA1 were selected for escaping from herd immunity (Koel et al., 2015; Lee et al., 2015). Notably, in both HA1 and HA2, π_N/π_S for 162− sequences was greater than that for 162+ sequences. This was consistent with the observation that π_N/π_S in the pandemic period was greater than that in the post-pandemic period for most of the proteins encoded by A(H1N1)pdm09, possibly due to relaxation of functional constraint or adaptation to new host

immediately after interspecies transmission (Su et al., 2015).

When the amino acid sites in HA1 and HA2 were categorized into solvent accessible and inaccessible sites (Supplementary Table S3), π_N/π_S was not significantly different between these categories in both HA1 and HA2 for either 162− or 162+ sequences (Fig. 1; Supplementary Table S4). The amino acid sites in HA1 were also categorized into those located inside or outside 10 Å or 15 Å from position 162. However, the number of sites located inside 10 Å appeared to be too small for obtaining reliable estimates of π_N/π_S (Supplementary Table S3). Therefore, π_N/π_S was computed for the sites located inside and outside 15 Å from position 162. In 162− sequences, π_N/π_S for the sites located inside 15 Å was not significantly smaller than one and was greater than that for the sites located outside 15 Å ($P < .05$ by Z-test) (Fig. 1; Supplementary Table S4). For both the sites located inside and outside 15 Å, π_N/π_S in 162+ sequences was smaller than that in 162− sequences ($P < .05$ by Z-test). In addition, difference in π_N/π_S between the sites located inside and outside 15 Å was obscured in 162+ sequences. Similar tendency was observed when the sites located inside and outside 15 Å were categorized into solvent accessible and inaccessible sites (Fig. 1; Supplementary Tables S3 and S4).

In the HA tree, 7562 162− and 1785 162+ sequences for A(H1N1)pdm09 made distinct clusters, although the bootstrap probability for the interior branch supporting the cluster of 162+ sequences was not significantly high (66%), as described above (Supplementary Fig. S1). However, in the NA tree constructed for the same set of strains as in the HA tree, it was observed that NA sequences for the strains with 162+ sequences of HA were included in the cluster of NA sequences for the strains with 162− sequences of HA, and *vice versa*, indicating occurrences of reassortments, although the bootstrap probabilities for the interior branches were again not significantly high (0–73%) (Supplementary Fig. S3) (Rambaut, 2018).

In the present study, the effect of *N*-linked glycosylation at position 162 of HA in A(H1N1)pdm09 was examined from molecular evolutionary analyses of nucleotide sequences. During evolution of A(H1N1)pdm09, the number of sequons increased in HA but not in NA, probably because humoral immunity was directed mainly against HA. In 162− sequences of HA, π_N/π_S for the amino acid sites located inside 15 Å from position 162 was relatively large irrespective of whether the sites were solvent accessible or inaccessible, suggesting that mutations at inaccessible sites as well as accessible sites may contribute to escaping from humoral immunity through epistatic effects (Kryazhimskiy et al., 2011). In 162+ sequences of HA, however, π_N/π_S for the sites located inside 15 Å from position 162 was small and comparable to that for the sites located outside 15 Å, suggesting that *N*-linked glycosylation at position 162 may shield epitopes against antibodies in A(H1N1)pdm09. It should be noted that the receptor binding pocket of HA in A(H1N1)pdm09 is composed of 190-helix (positions 184–191), 220-loop (positions 218–225), and 130-loop (positions 131–135), together with positions 91, 150, 180, and 192 forming the base (Yang et al., 2010). Since two (positions 184 and 218) of these positions were located inside 15 Å from position 162, *N*-linked glycosylation at position 162 may also interfere with the sialic acid receptor-binding activity of HA in A(H1N1)pdm09. However, the net-charge increased in both HA and NA, which may compensate for the reduction in the sialic acid receptor-binding activity of HA caused by the increase in the number of sequons (Kobayashi and Suzuki, 2012a). Reassortment apparently contributed to distributing 162+ sequences of HA among A(H1N1)pdm09 strains.

Reportedly, *N*-linked glycosylation of HA may be deleterious for A(H1N1)pdm09, because *N*-linked glycans may be recognized by lectins of the innate immune system (Tate et al., 2014; York et al., 2018). However, it was also demonstrated that *N*-linked glycans can be added to HA without eliciting the innate immune response (Hansen and Holmskov, 2002; Medina et al., 2013). Rather, *N*-linked glycosylation of HA may be advantageous for A(H1N1)pdm09, because HA without *N*-linked glycans may induce stress in the endoplasmic reticulum and trigger innate and adaptive immune responses (Hrincius et al., 2015).

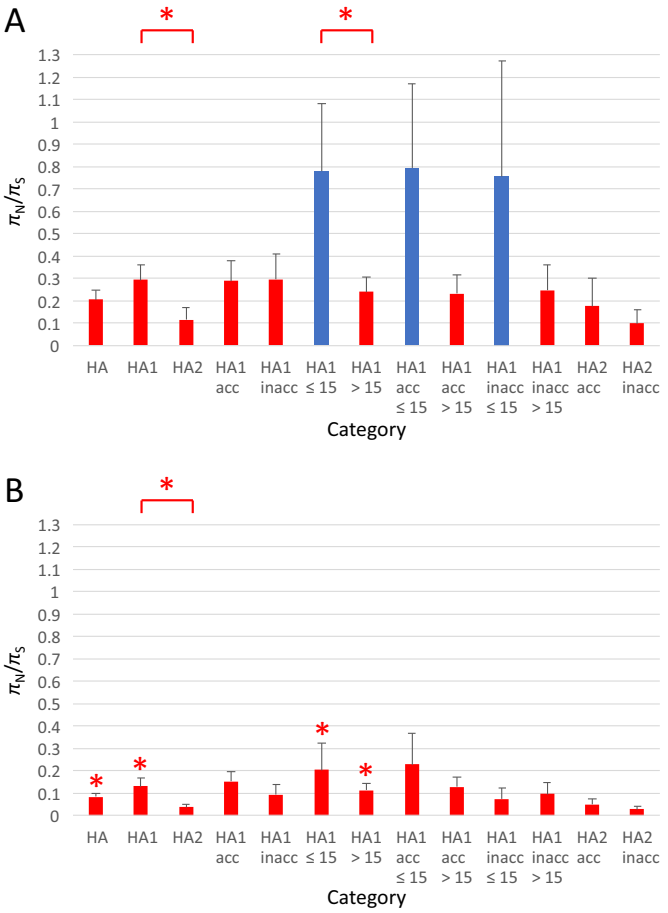

Fig. 1. Estimates of π_N/π_S with standard errors for structurally defined categories of amino acid sites in (A) 162− and (B) 162+ sequences of HA for A(H1N1)pdm09. The terms "acc" and "inacc" designate solvent accessible and inaccessible sites, respectively, and the terms "≤ 15" and "> 15" designate the sites whose α-carbons were located inside or outside 15 Å, respectively, from the α-carbon of the amino acid at position 162 in HA trimer for A(H1N1)pdm09 (PDB accession number: 4LXV). Red bars indicate that π_N was smaller than π_S with statistical significance ($P < .05$), whereas blue bars indicate that π_N and π_S were not different with statistical significance. The "*" signs placed between bars within (A) and (B) indicate that π_N/π_S was different between categories with statistical significance ($P < .05$). The "*" signs above bars in (B) indicate that π_N/π_S in (B) was smaller than the corresponding value in (A) with statistical significance ($P < .05$).

Notably, the sequon at positions 162–164 of HA changed from NXS to NXT during 2017–2018 in A(H1N1)pdm09, which was considered to facilitate *N*-linked glycosylation at position 162 (Altman et al., 2019). Taken together, these observations and the results obtained in the present study suggest that *N*-linked glycosylation at position 162 of HA may provide A(H1N1)pdm09 with selective advantage through shielding epitopes against antibodies.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.mgene.2020.100828.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

The authors thank five anonymous reviewers for valuable comments. This work was supported by JSPS KAKENHI Grant Number JP19K12221 and AMED Grant Number JP19fk0108033h0003 to Y.S.

References

Altman, M.O., Angel, M., Kosik, I., Trovao, N.S., Zost, S.J., Gibbs, J.S., Casalino, L., Amaro, R.E., Hensley, S.E., Nelson, M.I., Yewdell, J.W., 2019. Human influenza A virus hemagglutinin glycan evolution follows a temporal pattern to a glycan limit. mBio 10 (e00204-19).

Barr, I.G., Komadina, N., Hurt, A.C., Iannello, P., Tomasov, C., Shaw, R., Durrant, C., Sjogren, H., Hampson, A.W., 2005. An influenza A(H3) reassortant was epidemic in Australia and New Zealand in 2003. J. Med. Virol. 76, 391–397.

Chen, J., Sun, Y., 2011. Variation in the analysis of positively selected sites using nonsynonymous/synonymous rate ratios: an example using influenza virus. PLoS One 6, e19996.

Hansen, S., Holmskov, U., 2002. Lung surfactant protein D (SP-D) and the molecular diverted descendants: conglutinin, CL-43 and CL-46. Immunobiology 205, 498–517.

Hensley, S.E., Das, S.R., Bailey, A.L., Schmidt, L.M., Hickman, H.D., Jayaraman, A., Viswanathan, K., Raman, R., Sasisekharan, R., Bennink, J.R., Yewdell, J.W., 2009. Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift. Science 326, 734–736.

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.

Katoh, K., Misawa, K., Kuma, K.-I., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066.

Kobayashi, Y., Suzuki, Y., 2012a. Compensatory evolution of net-charge in influenza A virus hemagglutinin. PLoS One 7, e40422.

Kobayashi, Y., Suzuki, Y., 2012b. Evidence for N-glycan shielding of antigenic sites during evolution of human influenza A virus hemagglutinin. J. Virol. 86, 3446–3451.

Koel, B.F., Mogling, R., Chutinimitkul, S., Fraaij, P.L., Burke, D.F., van der Vliet, S., de Wit, E., Bestebroer, T.M., Rimmelzwaan, G.F., Osterhaus, A.D.M.E., Smith, D.J., Fouchier, R.A.M., de Graaf, M., 2015. Identification of amino acid substitutions supporting antigenic change of influenza A(H1N1)pdm09 viruses. J. Virol. 89, 3763–3775.

Krissinel, E., Henrick, K., 2007. Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 774–797.

Kryazhimskiy, S., Dushoff, J., Bazykin, G.A., Plotkin, J.B., 2011. Prevalence of epistasis in the evolution of influenza A surface proteins. PLoS Genet. 7, e1001301.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549.

Lee, A.J., Das, S.R., Wang, W., Fitzgerald, T., Pickett, B.E., Aevermann, B.D., Topham, D. J., Falsey, A.R., Scheuermann, R.H., 2015. Diversifying selection analysis predicts antigenic evolution of 2009 pandemic H1N1 influenza A virus in humans. J. Virol. 89, 5427–5440.

Lu, Y., Welsh, J.P., Swartz, J.R., 2014. Production and stabilization of the trimeric influenza hemagglutinin stem domain for potentially broadly protective influenza vaccines. Proc. Natl. Acad. Sci. U. S. A. 111, 125–130.

McAuley, J.L., Gilbertson, B.P., Trifkovic, S., Brown, L.E., McKimm-Breschkin, J.L., 2019. Influenza virus neuraminidase structure and functions. Front. Microbiol. 10, 39.

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., 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.

Mitnaul, L.J., Castrucci, M.R., Murti, K.G., Kawaoka, Y., 1996. The cytoplasmic tail of influenza A virus neuraminidase (NA) affects NA incorporation into virions, virion morphology, and virulence in mice but is not essential for virus replication. J. Virol. 70, 873–879.

Murrell, B., Wertheim, J.O., Moola, S., Weighill, T., Scheffler, K., Kosakovsky Pond, S.L., 2012. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 8, e1002764.

Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, Oxford, New York.

Nelson, M.I., Holmes, E.C., 2007. The evolution of epidemic influenza. Nat. Rev. Genet. 8, 196–205.

Ota, T., Nei, M., 1994. Variance and covariances of the numbers of synonymous and nonsynonymous substitutions per site. Mol. Biol. Evol. 11, 613–619.

Rambaut, A., 2018. FigTree. http://tree.bio.ed.ac.uk/software/figtree.

Skehel, J.J., Wiley, D.C., 2000. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annu. Rev. Biochem. 69, 531–569.

Su, Y.C.F., Bahl, J., Joseph, U., Butt, K.M., Peck, H.A., Koay, E.S.C., Oon, L.L.E., Barr, I.G., Vijaykrishna, D., Smith, G.J.D., 2015. Phylodynamics of H1N1/2009 influenza reveals the transition from host adaptation to immune-driven selection. Nat. Commun. 6, 7952.

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.

Suzuki, Y., 2006. Natural selection on the influenza virus genome. Mol. Biol. Evol. 23, 1902–1911.

Suzuki, Y., 2010. A phylogenetic approach to detecting reassortments in viruses with segmented genomes. Gene 464, 11–16.

Suzuki, Y., 2011. Positive selection for gains of N-linked glycosylation sites in hemagglutinin during evolution of H3N2 human influenza A virus. Genes Genet. Syst. 86, 287–294.

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.

de Vries, R.P., de Vries, E., Bosch, B.J., de Groot, R.J., Rottier, P.J.M., de Haan, C.A.M., 2010. The influenza A virus hemagglutinin glycosylation state affects receptor-binding specificity. Virology 403, 17–25.

Wagner, R., Matrosovich, M., Klenk, H.D., 2002. Functional balance between haemagglutinin and neuraminidase in influenza virus infections. Rev. Med. Virol. 12, 159–166.

Wang, C.-C., Chen, J.-R., Tseng, Y.-C., Hsu, C.-H., Hung, Y.-F., Chen, S.-W., Chen, C.-M., Khoo, K.-H., Cheng, T.-J., Cheng, Y.-S.E., Jan, J.-T., Wu, C.-Y., Ma, C., Wong, C.-H., 2009. Glycans on influenza hemagglutinin affect receptor binding and immune response. Proc. Natl. Acad. Sci. U. S. A. 106, 18137–18142.

Xu, R., Zhu, X., McBride, R., Nycholat, C.M., Yu, W., Paulson, J.C., Wilson, I.A., 2012. Functional balance of the hemagglutinin and neuraminidase activities accompanies the emergence of the 2009 H1N1 influenza pandemic. J. Virol. 86, 9221–9232.

Yan, L., Wang, H., Sun, L., Liu, Y., Sun, J., Zhao, X., Li, Y., Xie, X., Hu, J., 2019. An epitope on the stem region of hemagglutinin of H1N1 influenza A virus recognized by neutralizing monoclonal antibody. Biochem. Biophys. Res. Commun. 518, 319–324.

Yang, H., Carney, P., Stevens, J., 2010. Structure and receptor binding properties of a pandemic H1N1 virus hemagglutinin. PLoS Curr. 2, RRN1152.

York, I.A., Stevens, J., Alymova, I.V., 2018. Influenza virus N-linked glycosylation and innate immunity. Biosci. Rep. 39, BSR20171505.

Yoshida, R., Nei, M., 2016. Efficiencies of the NJp, maximum Likelihood, and Bayesian methods of phylogenetic construction for compositional and noncompositional genes. Mol. Biol. Evol. 33, 1618–1624.

Zhang, M., Gaschen, B., Blay, W., Foley, B., Haigwood, N., Kuiken, C., Bette 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.

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.