Methods and Applications

Mapping the Global Antigenic Evolution of Human Influenza A/H3N2 Neuraminidase Based on a Machine Learning Model — 1968–2024

Jingru Feng[1,&]; Rui Shi[2,&]; Huixin Zhou[3,&]; Shijie Wu[1,2]; Junyu Hu[1,2]; Taijiao Jiang[4,5,6]; Wenjie Han[2,#]; Xiangjun Du[1,3,7,#]

ABSTRACT

Introduction: Human influenza A/H3N2 imposes a substantial global disease burden. Beyond hemagglutinin (HA), neuraminidase (NA) also plays a critical role in the antigenic evolution of influenza viruses. However, a comprehensive understanding of NA antigenic evolution remains lacking.

Methods: NA inhibition (NAI) data were collected and structural epitopes for A/H3N2 NA were identified. A machine learning model was developed to accurately predict antigenic relationships by integrating four feature groups: epitopes, physicochemical properties, N-glycosylation, and catalytic sites. An antigenic correlation network (ACNet) was constructed and antigenic clusters were identified using the Markov clustering algorithm.

Results: The best random forest model (PREDEC-N2) achieved an accuracy of 0.904 in cross-validation and 0.867 in independent testing. Eight main antigenic clusters were identified on the ACNet. Spatiotemporal analysis revealed the continuous replacement and rapid global spread of new antigenic clusters for human influenza A/H3N2 NA.

Conclusions: This study developed a timely and accurate computational model to map the antigenic landscape of A/H3N2 NA, revealing both its relative antigenic conservation and continuous evolution. These insights provide valuable guidance for improved antigenic surveillance, vaccine recommendations, and prevention and control strategies for human influenza viruses.

Human influenza A/H3N2 has been a predominant seasonal influenza strain globally since its emergence in 1968. The main surface proteins of the influenza virus, hemagglutinin (HA) and neuraminidase (NA), evolve antigenically to escape immune recognition by the human host (*1*). Vaccination is the most effective intervention against influenza, but the vaccine effectiveness (VE) against H3N2 remains low (*2*). This low VE is mainly attributable to the rapid antigenic drift of HA and the insufficient induction of a robust NA-mediated immune response by current vaccines (*2*). Several studies have highlighted the critical role of NA-induced protection (*3*). Compared to human influenza A/H1N1, the antigenic divergence of NA in A/H3N2 is minimal but antigenic changes still occur (*4*). However, the antigenic evolution and landscape of NA in human influenza A/H3N2 remain poorly understood. To address this gap, we developed an antigenic classification model for human influenza A/H3N2 NA and identified distinct antigenic clusters to provide a more systematic understanding of the antigenic evolution of NA.

METHODS

Sequence Data

NA sequences of human influenza A/H3N2 viruses, available up to October 2024, were downloaded from global initiative on sharing all influenza data (GISAID) (*5*). To mitigate sampling bias, we implemented an even sampling strategy. Seven representative sequences were randomly selected for each month and each continent; if fewer than seven sequences were available, all sequences were included. Sequences containing more than three ambiguous amino acids or fewer than 400 residues in length were excluded. Subsequently, multiple sequence alignment was conducted, and three sequences with insertion mutations present in fewer than 1% of the sampled sequences were removed. The alignment was then repeated, resulting in a final sequence length of 469 amino acid residues. Sequence alignment and phylogenetic tree construction were performed using methods described in previous studies

(*6*). Finally, 9,054 sequences were analyzed.

A proportional sampling strategy was also implemented to avoid sampling error, selecting 5% of the sequences per month from each continent, or one sequence if the calculated sample size was less than one. After quality control, 7,847 sequences were retained for analysis.

NAI Data

A total of 376 pairs of NA inhibition (NAI) data were collected from various sources (*4,7–10*). For strain pairs tested in multiple experiments, the median result was used as the final value. The antigenic distance between two strains was calculated using the following formula (*11*):

$$H_{ab} = \sqrt{\frac{T_{aa}T_{bb}}{T_{ab}T_{ba}}} \tag{1}$$

where H_{ab} represents the NA antigenic distance between strain a and strain b, T_{ab} and T_{bb} are the NAI titers of serum b against virulent strains a and b, and T_{aa} and T_{ba} are the NAI titers of serum a against virulent strains a and b. A pair of strains was classified as antigenically similar if the absolute value of their antigenic distance was between 0.25 and 4 (not equal); otherwise, the pair was considered antigenically dissimilar.

Feature Selection

Twelve features were used to construct machine learning (ML) models based on NA sequences, which were categorized into four groups: epitopes, physicochemical properties, N-glycosylation, and catalytic sites.

Epitopes We used 7U4E as a template to identify potential structural epitopes. Sites with a binding probability above 0.1, as determined by ScanNet (*12*), were identified as potential epitope sites. K-means clustering was performed using spatial coordinates to determine the number of epitopes and composition of each epitope based on the Silhouette score (Supplementary Figure S1, available at https://weekly.chinacdc.cn/). Outliers that were excessively distant from other clusters were excluded, resulting in the identification of five epitopes (N2_A, N2_B, N2_C, N2_D, and N2_E, Supplementary Figure S2 and Supplementary Table S1, available at https://weekly.chinacdc.cn/). For each epitope, features were quantified by calculating the number of amino acid changes.

Physicochemical Properties Five classes of physicochemical properties were considered: hydrophobicity, charge, polarity, volume, and accessible surface area (ASA). A random forest (RF) model was trained on the training dataset to identify the best representative feature for each class. The selected indices were CHAM830107, RADA880108, CIDH920101, CHOC760102, and COHE430101. Features were computed by averaging the absolute differences between sequence pairs based on up to the three most prominent changes.

N-Glycosylation sites (N-Glycosylation) N-Glycosylation sites were identified using NetNGlyc (*13*), and the numbers of different glycosylation sites were calculated.

Catalytic sites (Catalyze) Eight previously reported NA catalytic sites were included, and the average Euclidean distances to the catalytic sites were calculated for each amino acid position, from which the three shortest distances were selected (*14*).

Model Construction

The antigenically similar or dissimilar label were used to train the model based on the 12-bit features calculated above. Five ML models capable of handling non-linear data were constructed using the Python package scikit-learn and evaluated: logistic regression (LR), support vector machine (SVM), K-nearest neighbors (KNN), RF, and extreme gradient boosting (XGBoost). We randomly split 70% of the NAI pairs for the training set and reserved the remaining 30% for the testing set. Model parameters were optimized using 5-fold cross-validation combined with random search conducted 500 times on the training set. The models were evaluated using five metrics: accuracy, precision, F1-score, recall, and receiver operating characteristic area under the curve (ROC-AUC).

Construction of Antigenic Network

The antigenic correlation network (ACNet) was constructed and visualized using Cytoscape (version 3.10.2, developed by Cytoscape Consortium, San Diego, United States). In this network, nodes represent NA strains and edges indicate antigenic similarity relationships as predicted by the model. The Markov cluster algorithm was used to identify clusters of strains based on the logarithmic ratio of the probabilities of antigenic similarity to dissimilarity. Clustering parameters were selected by optimizing mean cluster sizes and modularity (Supplementary Figure S3, available at https://weekly.chinacdc.cn/).

RESULTS

Antigenic Prediction Model for NA of Human Influenza A/H3N2

Five ML models were constructed using cross-validation (Supplementary Figure S4, available at https://weekly.chinacdc.cn/) and evaluated using the testing set (Figure 1A). The RF model outperformed all other models across all metrics, achieving the highest ROC-AUC value of 0.849 and the highest accuracy of 0.867 on the test set. Therefore, the RF

model was selected for subsequent analyses. Analysis of feature contributions revealed varying importance among different features (Figure 1B and C). Physicochemical properties contributed the most (39.1%), followed by epitopes (32.7%), catalytic sites (21.7%), and N-glycosylation sites (6.5%). The catalytic sites feature had the greatest individual impact, contributing approximately 21.7%. Among epitope-related features, Epitopes N2_D, N2_B, and N2_C, which are located near catalytic sites (Supplementary Figure S2), were identified as the most significant, indicating their critical role in both viral
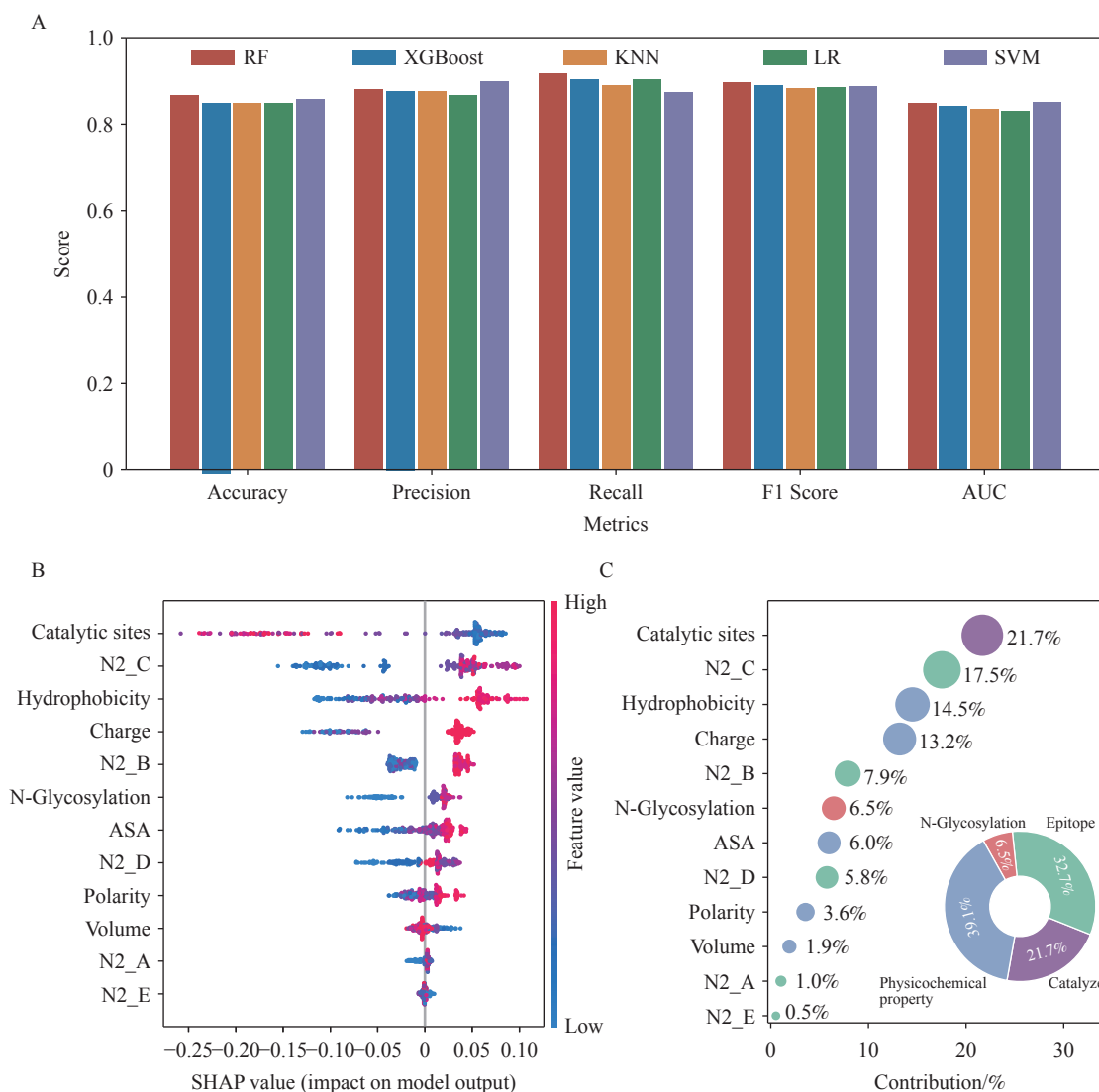

FIGURE 1. Model performance and feature contributions. (A) Model performance on the independent testing set; (B) Feature contributions at the sample level, where color indicates the magnitude and position reflects the absolute contribution of each feature; (C) Feature contributions at the population level.
Note: Different groups were color-coded, and size represents the magnitude of individual features. Abbreviation: RF=random forest; XGBoost=extreme gradient boosting; KNN=K-nearest neighbors; LR=logistic regression; SVM=support vector machine; AUC=receiver operating characteristic area under the curve; ASA=access surface area; SHAP=SHapley Additive exPlanation.

function and prediction (Figure 1C). The epitope located at the junction of different chains (N2_A) was less important, contributing only 1% to the overall feature importance.

The Antigenic Landscape for NA of Human Influenza A/H3N2

Based on the RF model, we predicted antigenic relationships between all representative strains for NA of human influenza A/H3N2 and constructed an ACNet for evenly sampled sequences. We identified eight major antigenic clusters that aligned with the traditional phylogenetic tree (Figure 2A). These clusters, which included vaccine strains, were named after the earliest vaccine strain within each cluster: PC73, TE77, BJ89, WH95, MS99, CA04, PE09, and SI16 (Supplementary Table S2, available at https://weekly.chinacdc.cn/). We validated the clustering by demonstrating that strains within the same cluster were more antigenically similar than those from different clusters (Supplementary Figure S5, available at https://weekly.chinacdc.cn/). A clear spatiotemporal pattern emerged: new clusters appeared and gradually replaced older ones, a trend observed consistently across different continents (Figure 2B and Supplementary Figure S6, available at https://weekly.chinacdc.cn/). Furthermore, NA antigenic clusters exhibited greater persistence over time (approximately 8 years) compared with HA (approximately 2 or 3 years) (Figure 2C) (15). The clustering and prevalence analyses from proportional sampling were largely consistent with these findings (Supplementary Figure S7, available at https://weekly.chinacdc.cn/).

DISCUSSION

In the present study, we developed a novel machine learning model for timely and effective prediction of antigenic relationships in the neuraminidase of human influenza A/H3N2. We identified eight main antigenic clusters between 1968 and 2024. Spatiotemporal analysis revealed continuous global replacement and rapid spread of new antigenic clusters. Our findings were robust across different sampling approaches. Among forty-eight vaccine strains, only one (A/Wellington/01/2004) during the cluster transition period was classified differently, likely due to significant differences in sequence distribution.

The antigenic prediction model for NA was developed using an approach similar to that for HA. A key adjustment was replacing the receptor-binding features of HA with catalytic sites, which are more pertinent to NA function. Additionally, we identified NA epitopes de novo for feature calculation. While these adjustments did not represent significant innovations, the framework has proven effective with only minor modifications across different contexts (16). This suggests that with appropriate adjustments, our model can provide accurate predictions for NA antigenic correlations.

Both antigenic clusters and phylogenetic clades reflect evolutionary relationships between viral strains, representing phenotype and genotype, respectively. Unlike the continuous branching of phylogeny, antigenic clusters represent important discrete phenotypes for HA and NA, with nonlinear relationships to genetic changes (4). Variations at different sites have inconsistent effects on antigenicity. The phylogenetic tree for NA displayed a single-trunk structure, indicating minimal selection pressure. Furthermore, spatiotemporal analysis confirmed the continuous global replacement of older antigenic clusters by newer ones. Only eight major antigenic clusters were identified for NA over the past 60 years (approximately one cluster every 8 years), significantly fewer than for HA (approximately every 2 or 3 years). This phenomenon might be explained by the relatively lower mean rate of nucleotide substitution in NA, which could be partly attributed to stronger structural constraints on this enzyme compared to the receptor-binding protein, as well as the stronger selection pressure and greater immune pressure on HA, likely due to its role as the primary vaccine target and its higher distribution on the virus surface (17–18).

This study has several limitations that warrant consideration. First, while our dataset was sufficient for model development, a larger dataset would enable the construction of more sophisticated models with improved prediction performance. Second, although we developed the first ML antigenic classification model for NA, integration with HA and other important viral components is necessary for a comprehensive understanding of antigenic evolution and its implications for seasonal influenza. Third, due to variations in sequencing coverage, some regions had insufficient sequence data, leading to incomplete characterization of their antigenic landscapes. Similarly, the analysis of early antigenic clusters may be subject to sequencing bias. Finally, while our model demonstrated high predictive accuracy, validation with experimental data or real-world outcomes would further strengthen its applicability.

The findings of this study highlight the crucial role of NA in the antigenic evolution of human influenza A/H3N2 and its contribution to viral circulation and
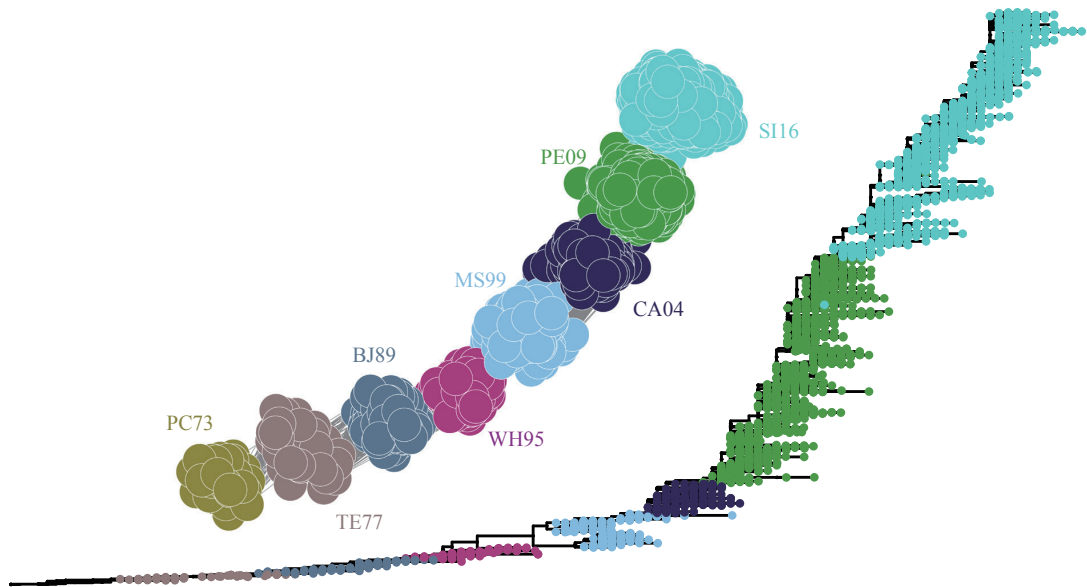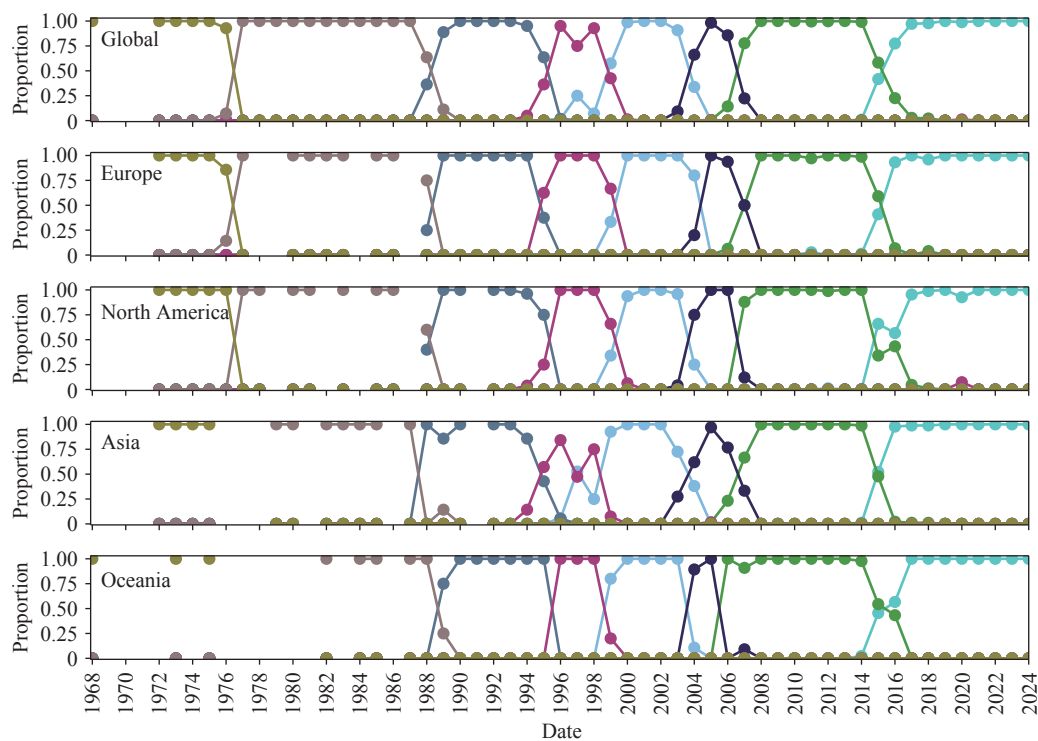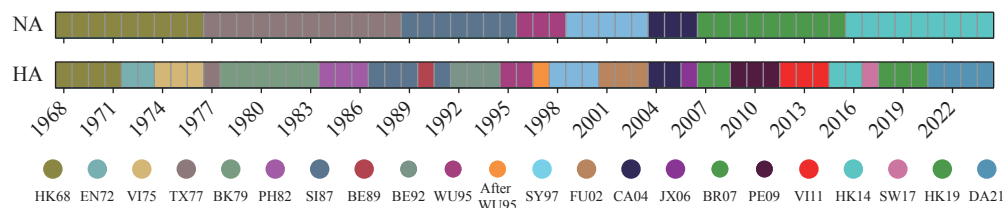
FIGURE 2. Antigenic landscape of A/H3N2 NA. (A) The ACNet and phylogenetic tree for representative sequences from eight major antigenic clusters; (B) Yearly spatiotemporal distribution of eight antigenic clusters; (C) Replacement patterns of dominant antigenic clusters.

Note: The legend displays colors for HA while the color of NA antigenic clusters correspond to those used in the ACNet. For (C), color changes indicating that a new antigenic cluster is dominant or becoming dominant.

Abbreviation: NA=neuraminidase; HA=hemagglutinin; ACNet=antigenic correlation network.

spread. Although NA currently receives less consideration in vaccine strain recommendation and antigenic surveillance, the tools developed in this study can facilitate improved antigenic monitoring, inform vaccine selection, and ultimately aid in the prevention and control of influenza epidemics as knowledge deepens and relevant technologies advance.

Conflicts of interest: No conflicts of interest.

Corresponding authors: Wenjie Han, hanwj7@mail2.sysu.edu.cn; Xiangjun Du, duxj9@mail.sysu.edu.cn.

1 School of Public Health (Shenzhen), Sun Yat-sen University, Guangzhou City, Guangdong Province, China; 2 School of Public Health (Shenzhen), Shenzhen Campus of Sun Yat-sen University, Shenzhen City, Guangdong Province, China; 3 Shenzhen Key Laboratory of Pathogenic Microbes & Biosafety, Shenzhen Campus of Sun Yat-sen University, Shenzhen City, Guangdong Province, China; 4 Guangzhou National Laboratory, Guangzhou City, Guangdong Province, China; 5 State Key Laboratory of Respiratory Disease, The Key Laboratory of Advanced Interdisciplinary Studies Center, the First Affiliated Hospital of Guangzhou Medical University, Guangzhou City, Guangdong Province, China; 6 Suzhou Institute of Systems Medicine, Suzhou City, Jiangsu Province, China; 7 Key Laboratory of Tropical Disease Control, Ministry of Education, Sun Yat-sen University, Guangzhou City, Guangdong Province, China.
& Joint first authors.

REFERENCES

1. Petrova VN, Russell CA. The evolution of seasonal influenza viruses. Nat Rev Microbiol 2018;16(1):47 – 60. https://doi.org/10.1038/nrmicro.2017.118.
2. Chen YQ, Wohlbold TJ, Zheng NY, Huang M, Huang YP, Neu KE, et al. Influenza infection in humans induces broadly cross-reactive and protective neuraminidase-reactive antibodies. Cell 2018;173(2):417 – 29.e10. https://doi.org/10.1016/j.cell.2018.03.030.
3. Weiss CD, Wang W, Lu Y, Billings M, Eick-Cost A, Couzens L, et al. Neutralizing and neuraminidase antibodies correlate with protection against influenza during a late season A/H3N2 outbreak among unvaccinated military recruits. Clin Infect Dis 2020;71(12):3096 – 102. https://doi.org/10.1093/cid/ciz1198.
4. Sandbulte MR, Westgeest KB, Gao J, Xu XY, Klimov AI, Russell CA, et al. Discordant antigenic drift of neuraminidase and hemagglutinin in H1N1 and H3N2 influenza viruses. Proc Natl Acad Sci USA 2011;108 (51):20748 – 53. https://doi.org/10.1073/pnas.1113801108.
5. Khare S, Gurry C, Freitas L, Schultz MB, Bach G, Diallo A, et al. GISAID's role in pandemic response. China CDC Wkly 2021;3(49): 1049 – 51. https://doi.org/10.46234/ccdcw2021.255.
6. Zhai K, Dong JZ, Zeng JF, Cheng PW, Wu XS, Han WJ, et al. Global antigenic landscape and vaccine recommendation strategy for low pathogenic avian influenza A (H9N2) viruses. J Infect 2024;89(2): 106199. https://doi.org/10.1016/j.jinf.2024.106199.
7. Catani JPP, Smet A, Ysenbaert T, Vuylsteke M, Bottu G, Mathys J, et al. The antigenic landscape of human influenza N2 neuraminidases from 2009 until 2017. eLife 2024;12:RP90782 https://doi.org/10.7554/eLife.90782.4.
8. Gao J, Li X, Klenow L, Malik T, Wan HQ, Ye ZP, et al. Antigenic comparison of the neuraminidases from recent influenza A vaccine viruses and 2019–2020 circulating strains. npj Vaccines 2022;7(1):79 https://doi.org/10.1038/s41541-022-00500-1.
9. Schild GC, Oxford JS, Dowdle WR, Coleman M, Pereira MS, Chakraverty P. Antigenic variation in current influenza A viruses: evidence for a high frequency of antigenic 'drift' for the Hong Kong virus. Bull World Health Organ 1974;51(1):1-11. https://pubmed.ncbi.nlm.nih.gov/4218138/.
10. Kilbourne ED, Johansson BE, Grajower B. Independent and disparate evolution in nature of influenza A virus hemagglutinin and neuraminidase glycoproteins. Proc Natl Acad Sci USA 1990;87(2): 786 – 90 https://doi.org/10.1073/pnas.87.2.786.
11. Meng J, Liu JZ, Song WK, Li HL, Wang JY, Zhang L, et al. PREDAC-CNN: predicting antigenic clusters of seasonal influenza A viruses with convolutional neural network. Brief Bioinform 2024;25(2):bbae033. https://doi.org/10.1093/bib/bbae033.
12. Tubiana J, Schneidman-Duhovny D, Wolfson HJ. ScanNet: an interpretable geometric deep learning model for structure-based protein binding site prediction. Nat Methods 2022;19(6):730 – 9. https://doi.org/10.1038/s41592-022-01490-7.
13. Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S, Brunak S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics 2004;4(6):1633 – 49. https://doi.org/10.1002/pmic.200300771.
14. Jagadesh A, Salam AAA, Mudgal PP, Arunkumar G. Influenza virus neuraminidase (NA): a target for antivirals and vaccines. Arch Virol 2016;161(8):2087 – 94. https://doi.org/10.1007/s00705-016-2907-7.
15. Du XJ, Dong LB, Lan Y, Peng YS, Wu AP, Zhang Y, et al. Mapping of H3N2 influenza antigenic evolution in China reveals a strategy for vaccine strain recommendation. Nat Commun 2012;3(1):709. https://doi.org/10.1038/ncomms1710.
16. Qiu JX, Qiu TY, Dong QL, Xu DP, Wang X, Zhang Q, et al. Predicting the antigenic relationship of foot-and-mouth disease virus for vaccine selection through a computational model. IEEE/ACM Trans Comput Biol Bioinf 2021;18(2):677 – 85. https://doi.org/10.1109/TCBB.2019.2923396.
17. Westgeest KB, de Graaf M, Fourment M, Bestebroer TM, van Beek R, Spronken MIJ, et al. Genetic evolution of the neuraminidase of influenza A (H3N2) viruses from 1968 to 2009 and its correspondence to haemagglutinin evolution. J Gen Virol 2012;93(9):1996 – 2007. https://doi.org/10.1099/vir.0.043059-0.
18. Harris A, Cardone G, Winkler DC, Heymann JB, Brecher M, White JM, et al. Influenza virus pleiomorphy characterized by cryoelectron tomography. Proc Natl Acad Sci USA 2006;103(50):19123 – 7. https://doi.org/10.1073/pnas.0607614103.

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY TABLE 1. NA inhibition experimental data (*1–5*).

The complete table is available at: https://zenodo.org/records/15653659 (uploaded on June 13, 2025).

SUPPLEMENTARY TABLE 2. Identified epitopes for human influenza A/H3N2 NA.

Epitopes	Sites	Outliers
N2_A	82, 83, 84, 86, 88, 89, 90, 187, 207, 208, 234, 236, 258, 259, 283, 284, 285, 286, 306, 307, 308, 309, 311, 357, 415, 416	–
N2_B	118, 143, 146, 147, 150, 151, 152, 153, 154, 156, 178, 406, 430, 431, 432, 433, 434, 437, 469	–
N2_C	195, 196, 197, 198, 199, 200, 218, 219, 220, 221, 222, 224, 244, 245, 246, 247, 248, 249, 250, 251, 253, 268, 273, 274, 276, 277	–
N2_D	292, 294, 295, 296, 326, 327, 328, 329, 331, 332, 333, 334, 339, 342, 343, 344, 345, 346, 347, 348, 358, 369, 370, 371, 384, 385, 390	358,384,385,390
N2_E	392, 394, 399, 400, 401, 402, 453, 454, 455, 456, 457, 458	–

Note: "–" means no outlier sites were identified for the corresponding epitope.
Abbreviation: NA=neuraminidase.

SUPPLEMENTARY TABLE 3. Vaccine strains included in each antigenic cluster.

Vac_strains	Accession number	Cluster
A/Port_chalmers/1/1973	EPI_ISL_20940	PC73
A/Victoria/3/1975	EPI_ISL_113968	PC73
A/Texas/1/1977	EPI_ISL_122019	TE77
A/Bangkok/1/1979	EPI_ISL_122020	TE77
A/Philippines/2/1982	EPI_ISL_76487	TE77
A/Leningrad/360/1986	EPI_ISL_124928	TE77
A/Shanghai/11/1987	EPI_ISL_114294	TE77
A/Sichuan/2/1987	EPI_ISL_125877	TE77
A/Beijing/353/1989	EPI_ISL_115769	BJ89
A/Guizhou/54/1989	EPI_ISL_124894	BJ89
A/Beijing/32/1992	EPI_ISL_22624	BJ89
A/Shangdong/9/1993	EPI_ISL_115410	BJ89
A/Johannesurg/33/1994	EPI_ISL_14284833	BJ89
A/Wuhan/359/1995	EPI_ISL_111269	WH95
A/Nanchang/933/1995	EPI_ISL_14284834	WH95
A/Sydney/5/1997	EPI_ISL_14284880	WH95
A/Moscow/10/1999	EPI_ISL_127595	MS99
A/Panama/2007/1999	EPI_ISL_174195	WH95
A/Fujian/411/2002	EPI_ISL_107711	MS99
A/Wyoming/03/2003	EPI_ISL_153570	MS99
A/Fujian/445/2003	EPI_ISL_127602	MS99
A/California/7/2004	EPI_ISL_21094	CA04
A/Wellington/01/2004	EPI_ISL_127603	CA04*
A/Newyork/55/2004	EPI_ISL_2402569	CA04
A/Hiroshima/52/2005	EPI_ISL_13228	CA04
A/Wisconsin/67/2005	EPI_ISL_10430	CA04
A/Jiangxi/donghu312/2006	EPI_ISL_20194	CA04

Continued

| Vac_strains	Accession number	Cluster
A/Brisbane/10/2007	EPI_ISL_110723	PE09
A/Perth/16/2009	EPI_ISL_31913	PE09
A/Fujian-tongan/196/2009	EPI_ISL_99110	PE09
A/Victoria/361/2011	EPI_ISL_109762	PE09
A/Texas/50/2012	EPI_ISL_122006	PE09
A/Switzerland/9715293/2013	EPI_ISL_166859	PE09
A/Hong_kong/4801/2014	EPI_ISL_176512	PE09
A/Singapore/infimh-16-0019/2016	EPI_ISL_330262	SI16
A/Kansas/14/2017	EPI_ISL_312833	SI16
A/Switzerland/8060/2017	EPI_ISL_303951	SI16
A/Southaustralia/34/2019	EPI_ISL_17668123	SI16
A/Hong_kong/2671/2019	EPI_ISL_391201	SI16
A/Hongkong/45/2019	EPI_ISL_410589	SI16
A/Cambodia/e0826360/2020	EPI_ISL_1367566	SI16
A/Darwin/6/2021	EPI_ISL_18879343	SI16
A/Darwin/9/2021	EPI_ISL_16998754	SI16
A/Thailand/8/2022	EPI_ISL_18485009	SI16
A/Massachusetts/18/2022	EPI_ISL_16968012	SI16
A/Croatia/10136rv/2023	EPI_ISL_19085873	SI16
A/District_of_ columbia/27/2023	EPI_ISL_19175844	SI16

* Derived from even sampling.
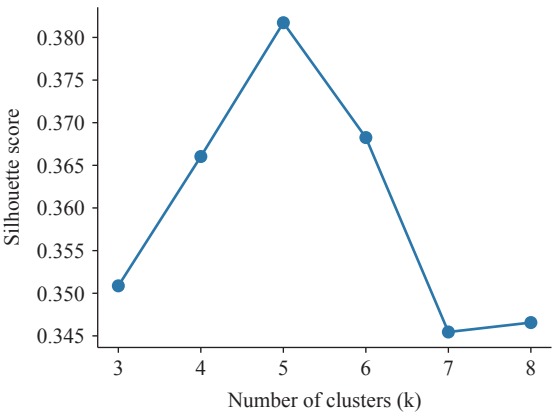
SUPPLEMENTARY FIGURE S1. Silhouette score curve for different values of k (number of epitopes).
Note: A higher score represents better clustering results.
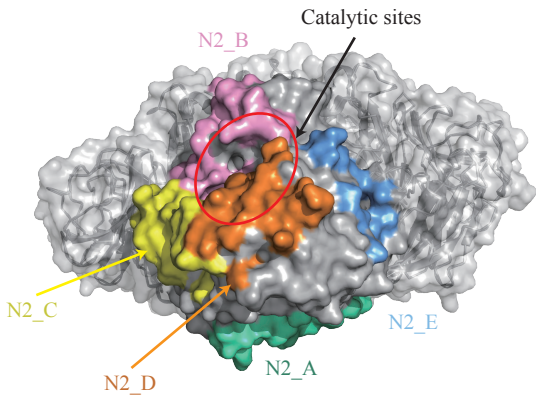
SUPPLEMENTARY FIGURE S2. Identified epitopes. Different colors indicate distinct epitopes.
Note: The red circle indicates catalytic sites. Epitopes are shown on only one chain of the tetramer.
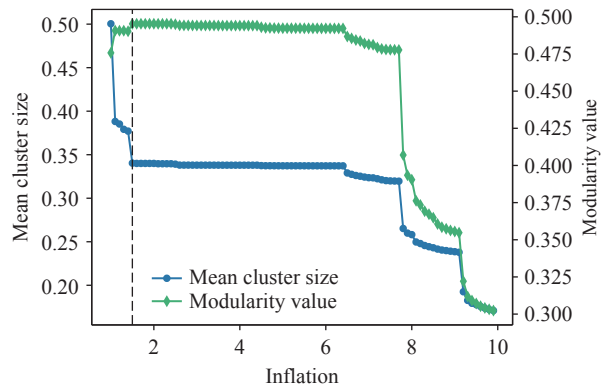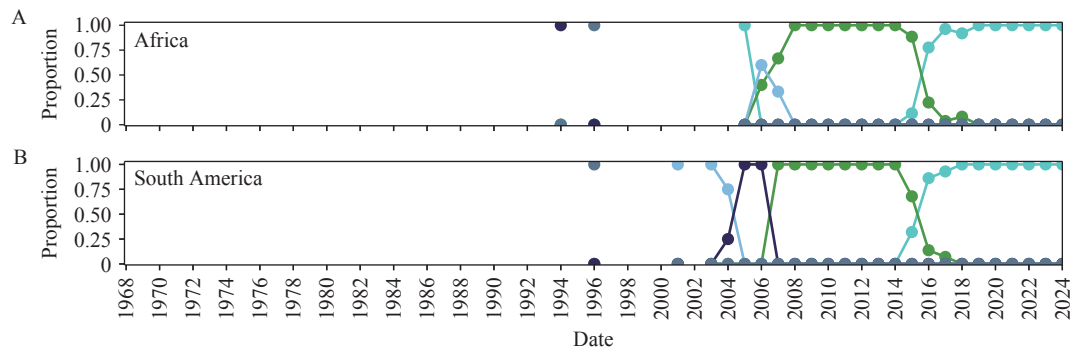
SUPPLEMENTARY FIGURE S3. Average cluster size and modularity curve.
Note: The blue line represents the mean cluster size (left y-axis). The green line represents modularity (right y-axis). The parameter corresponding to the first point in the first plateau period was selected, with higher modularity representing better clustering results. An inflation value of 1.5 for MCL was selected.
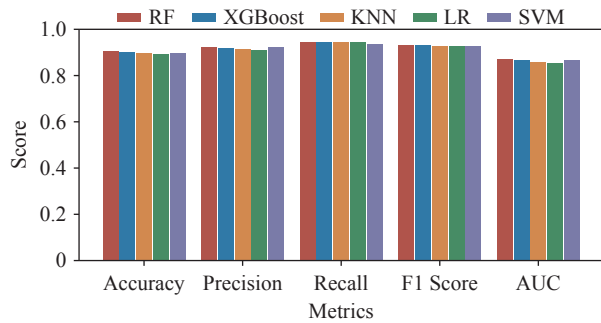
SUPPLEMENTARY FIGURE S4. Performance of different models on the training dataset based on 5-fold cross-validation.
Abbreviation: RF=random forest; XGBoost=extreme gradient boosting; KNN=K-nearest neighbors; LR=logistic regression; SVM=support vector machine; AUC=receiver operating characteristic area under the curve.
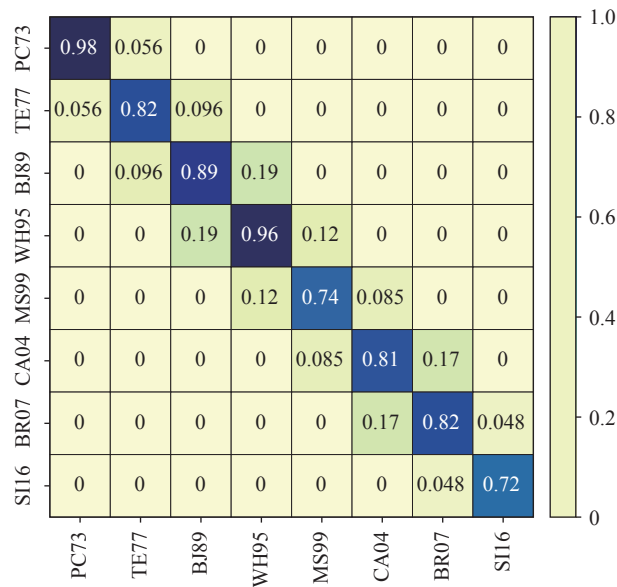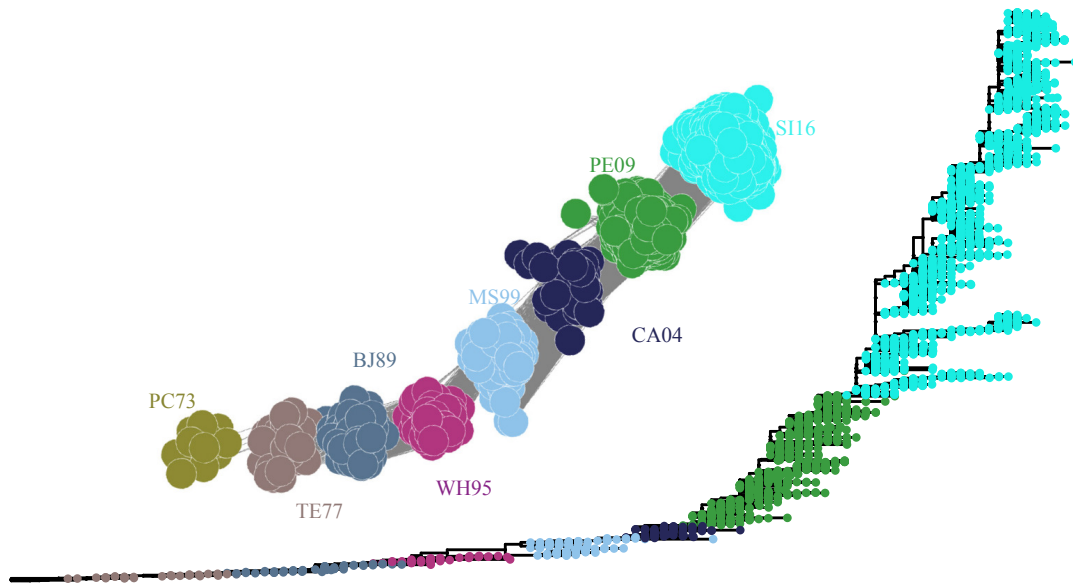

SUPPLEMENTARY FIGURE S5. Antigenic similarity proportion for eight antigenic clusters.
Note: The proportions were calculated based on all possible pairs within a cluster and between two clusters.
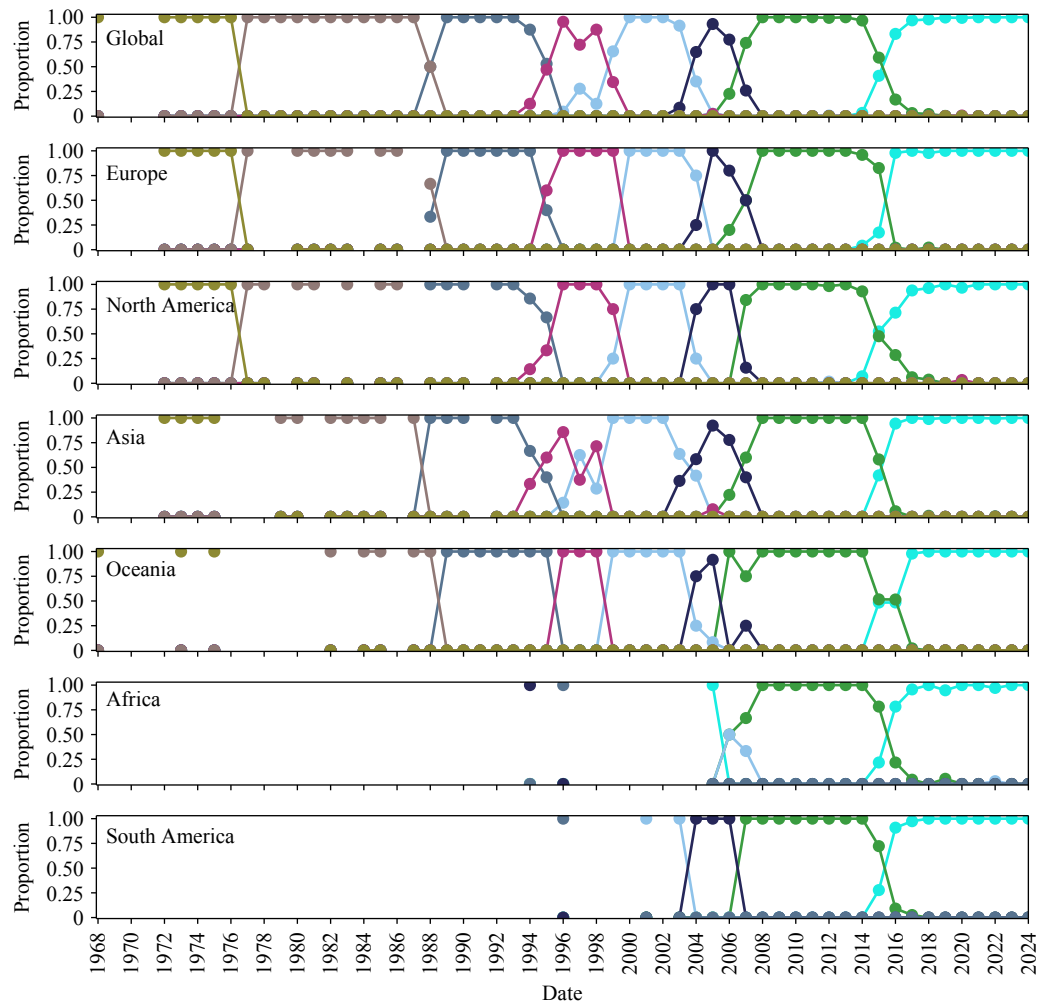
SUPPLEMENTARY FIGURE S6. Spatiotemporal distribution of even sampling for eight antigenic clusters in (A) Africa, and (B) South America.

SUPPLEMENTARY FIGURE S7. Antigenic landscape of A/H3N2 NA from proportional sampling. (A) The ACNet and phylogenetic tree of representative sequences from eight major antigenic clusters; (B) Yearly spatiotemporal distribution of eight antigenic clusters.
Abbreviation: NA=neuraminidase; ACNet=antigenic correlation network.

REFERENCES

1. Catani JPP, Smet A, Ysenbaert T, Vuylsteke M, Bottu G, Mathys J, et al. The antigenic landscape of human influenza N2 neuraminidases from 2009 until 2017. eLife 2024;12:RP90782. https://doi.org/10.7554/eLife.90782.4.
2. Sandbulte MR, Westgeest KB, Gao J, Xu XY, Klimov AI, Russell CA, et al. Discordant antigenic drift of neuraminidase and hemagglutinin in H1N1 and H3N2 influenza viruses. Proc Natl Acad Sci USA 2011;108(51):20748 – 53. https://doi.org/10.1073/pnas.1113801108.
3. Gao J, Li X, Klenow L, Malik T, Wan HQ, Ye ZP, et al. Antigenic comparison of the neuraminidases from recent influenza A vaccine viruses and 2019–2020 circulating strains. npj Vaccines 2022;7(1):79. https://doi.org/10.1038/s41541-022-00500-1.
4. Schild GC, Oxford JS, Dowdle WR, Coleman M, Pereira MS, Chakraverty P. Antigenic variation in current influenza A viruses: evidence for a high frequency of antigenic 'drift' for the Hong Kong virus. Bull World Health Organ 1974;51(1):1-11. https://pubmed.ncbi.nlm.nih.gov/4218138/.
5. Kilbourne ED, Johansson BE, Grajower B. Independent and disparate evolution in nature of influenza A virus hemagglutinin and neuraminidase glycoproteins. Proc Natl Acad Sci USA 1990;87(2):786 – 90. https://doi.org/10.1073/pnas.87.2.786.