Journal of Infection 89 (2024) 106199

Contents lists available at ScienceDirect

Journal of Infection

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

Viruses and Viral Diseases

Global antigenic landscape and vaccine recommendation strategy for low pathogenic avian influenza A (H9N2) viruses

Ke Zhai [a,b,1], Jinze Dong [c,1], Jinfeng Zeng [a,b], Peiwen Cheng [a,b], Xinsheng Wu [a,b], Wenjie Han [a,b], Yilin Chen [a,b], Zekai Qiu [a,b,d,e], Yong Zhou [c], Juan Pu [c,*], Taijiao Jiang [f,g,h,**], Xiangjun Du [a,b,i,j,***]

[a] School of Public Health (Shenzhen), Sun Yat-sen University, Guangzhou 510275, PR China
[b] School of Public Health (Shenzhen), Shenzhen Campus of Sun Yat-sen University, Shenzhen 518107, PR China
[c] National Key Laboratory of Veterinary Public Health and Safety, Key Laboratory for Prevention and Control of Avian Influenza and Other Major Poultry Diseases, Ministry of Agriculture and Rural Affairs, College of Veterinary Medicine, China Agricultural University, Beijing 100193, PR China
[d] Department of Molecular and Radiooncology, German Cancer Research Center (DKFZ), Heidelberg 69120, Germany
[e] Medical Faculty Heidelberg, Heidelberg University, Heidelberg 69047, Germany
[f] Guangzhou National Laboratory, Guangzhou 510005, PR China
[g] State Key Laboratory of Respiratory Disease, The Key Laboratory of Advanced Interdisciplinary Studies Center, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou 510120, PR China
[h] Suzhou Institute of Systems Medicine, Suzhou 215123, PR China
[i] Shenzhen Key Laboratory of Pathogenic Microbes & Biosecurity, Shenzhen Campus of Sun Yat-sen University, Shenzhen 518107, PR China
[j] Key Laboratory of Tropical Disease Control, Ministry of Education, Sun Yat-sen University, Guangzhou 510030, PR China

ARTICLE INFO

Article history:
Accepted 11 June 2024
Available online 18 June 2024

Keywords:
Avian influenza
H9N2
Antigenic cluster
Surveillance
Vaccine recommendation

SUMMARY

The sustained circulation of H9N2 avian influenza viruses (AIVs) poses a significant threat for contributing to a new pandemic. Given the temporal and spatial uncertainty in the antigenicity of H9N2 AIVs, the immune protection efficiency of vaccines remains challenging. By developing an antigenicity prediction method for H9N2 AIVs, named PREDAC-H9, the global antigenic landscape of H9N2 AIVs was mapped. PREDAC-H9 utilizes the XGBoost model with 14 well-designed features. The XGBoost model was built and evaluated to predict the antigenic relationship between any two viruses with high values of 81.1 %, 81.4 %, 81.3 %, 81.1 %, and 89.4 % in accuracy, precision, recall, F1 value, and area under curve (AUC), respectively. Then the antigenic correlation network (ACnet) was constructed based on the predicted antigenic relationship for H9N2 AIVs from 1966 to 2022, and ten major antigenic clusters were identified. Of these, four novel clusters were generated in China in the past decade, demonstrating the unique complex situation there. To help tackle this situation, we applied PREDAC-H9 to calculate the cluster-transition determining sites and screen out virus strains with the high cross-protective spectrum, thus providing an in silico reference for vaccine recommendation. The proposed model will reduce the clinical monitoring workload and provide a useful tool for surveillance and control of H9N2 AIVs.

Introduction

The H9N2 influenza virus is an important pathogen impacting both animal and human health. Since the H9N2 AIVs were first isolated in 1966, this subtype of low pathogenicity AIVs has been prevalent in various countries and regions globally over the subsequent decades.[1–3] Their most significant impact in chicken is a sudden decrease in appetite at a large scale, resulting in secondary bacterial infections and finally leading to egg drop up and economic losses of up to 30 %.[4] Notably, in China, H9N2 has replaced H5N1 and became the most dominant subtype in domestic poultry since 2016.[5]

As of September 14, 2023, there have been 127 reported human cases of low pathogenic avian influenza (LPAI) H9N2 worldwide,

* Correspondence to: College of Veterinary Medicine, China Agricultural University, Beijing 100193, PR China.
** Corresponding author at: Guangzhou National Laboratory, Guangzhou 510005, PR China.
*** Corresponding author at: School of Public Health (Shenzhen), Sun Yat-sen University, Guangzhou 510275, PR China.
E-mail addresses: pujuan@cau.edu.cn (J. Pu), jiang_taijiao@gzlab.ac.cn (T. Jiang), duxj9@mail.sysu.edu.cn (X. Du).
[1] These authors have contributed equally to the work.

https://doi.org/10.1016/j.jinf.2024.106199

characterized by flu-like symptoms.[6] Over the past three decades, more than 100 human infection cases have been recorded in China. Confirmed cases have shown a rising trend in recent years, with 14 from 1998 to 2012, 37 from 2013 to 2019, and 63 from 2020 to September 2023.[6] As the clinic symptoms of most H9N2 human infection cases were mild and self-limited, public as well as clinical physicians tend to pay relatively less attention compared with H5N1 and H7N9 human infections. However, serological evidence has shown a higher prevalence of H9N2 in poultry workers compared with H5N1 or H7N9 subtypes.[7,8] Additionally, H9N2 AIVs have played an important role in the evolution of AIVs by providing various internal genes required for gene reassortment, leading to the emergence of novel AIVs (including H5N1,[3] H5N6,[9] H7N9,[10,11] and so on). Therefore, H9N2 is listed by the WHO as one of the most likely pathogens to cause a pandemic.[12]

Antigenic shifts in influenza viruses pose a significant threat since they can result in novel viruses to which populations have no immunity.[13,14] Three of the four pandemic influenza outbreaks in the 20th century were the result of antigenic shift events that involved AIVs.[14,15] H9N2 AIVs have experienced multiple antigenic drifts and caused outbreaks in poultry flocks. In 2013, antigenically mutated strains of H9N2 viruses caused a large outbreak in Chinese chickens and created precursor conditions for the subsequent generation of H7N9 viruses.[16] In recent years, H9N2 continues to mutate and generate novel antigenic variants,[17–19] causing co-prevalence of multiple antigenic strains.[20,21] Although vaccination is currently a commonly used prevention and control strategy in poultry farms, the protective efficiency of vaccines continues to be challenged as the antigenicity of the virus evolves. Effective prevention strategies against H9N2 AIVs, such as rapid production of matching vaccine strains and expanded vaccine protection range, are urgently needed.[2]

To help achieve this strategy, it is crucial to understand the antigenic evolution pattern of the H9N2 AIVs and to identify the newly emerging antigenic clusters. However, the traditional method of determining influenza antigenicity through hemagglutination inhibition (HI) assay is impractical for large-scale screening of AIV strains due to their time-consuming, labor-intensive, and demanding experimental conditions. In this study, we developed an antigenic relationship machine-learning prediction model based on H9N2 HA sequencing data and HI data. This allowed us to rapidly infer the antigenic pattern of H9N2 AIVs and improve vaccine strain recommendations.

Materials and methods

HA sequence and structure data

The HA1 protein sequence of H9N2 AIV, the main immunogen of influenza virus, was obtained from the NCBI Influenza Virus Resource and the Global Initiative on Sharing All Influenza Data (GISAID),[22] covering the period from 1966 to 2022. The HA1 dataset comprised a total of 10,289 sequences, with each sequence consisting of 317 amino acids. Multiple sequence alignment was performed using the MAFFT,[23] and sequences with a proportion of missing or aberrant amino acids exceeding 10 % (i.e., "-", "X") in the HA1 domain were excluded. Subsequently, the missing or aberrant amino acids in the remaining sequences were filled. Specifically, these positions were filled with the mode of the amino acid types from the top ten sequences that exhibited the highest similarity to the target sequence.

Furthermore, the structural data for the H9N2 HA monomer, 1JSD (strain A/Swine/Hong Kong/9/1998), was downloaded from the Protein Data Bank (PDB) database.[24] The structure data of H9N2 HA trimer, P03457 (strain A/Turkey/Wisconsin/1/1966), was obtained from the SWISS-MODEL website.[25]

HI data for H9N2 AIV

We collected and combined datasets of HI measurements from various literature sources, and standardized log-transformed titers[26] were used to measure the antigenic distance between viruses A and B. The distance is defined as follows:

$$dHI = \log_2\left(\frac{T_{b\beta}}{T_{a\beta}}\right) \quad (1)$$

Here, dHI is standardized HI titer of a test virus a using antiserum β raised against the reference virus b. $T_{b\beta}$ represents the homologous titer of virus b detected by antiserum β. $T_{a\beta}$ represents the heterologous titer of virus a detected by antiserum β. A pair of viruses were considered antigenically different if $|dHI| \geq 2$, otherwise, they were considered antigenically similar.[27] Considering the different experimental conditions in different laboratories, we took the median titer results for the same pair of strains in different HI test tables.[28] Overall, we obtained 2388 (57.3 %) antigenic variant pairs and 1773 (42.6 %) antigenic similar pairs.

Extension of antigenic sites of H9N2 AIV

Like human influenza viruses, avian H9N2 viruses mainly change the major antigenic sites of their HA proteins to evade antibody attacks. To predict the protein-antibody binding sites, we utilized the Spatio-Chemical Arrangement of Neighbors Network (ScanNet) webserver,[29,30] which provides a high level of accuracy, speed, and coverage compared to previous approaches. Firstly, we utilized the P03457 trimer protein structure[25] as input and selected the top 150 residues for each monomer based on the probability of binding sites extracted from the output file. Secondly, we took the intersection of the residue positions from the three monomers. Thirdly, we manually checked the position of residues on the protein structure and removed stem residues. Fourthly, previous studies have defined four potential antigenic epitopes for the H9N2-subtype hemagglutinin: Site I, Site II, H9-A, and H9-B, collectively comprising a total of 26 residues.[31–34] Of these 26 residues, 23 were included in the previous step, and we added the remaining three to obtain a final count of 96 residues (Fig. S3). Lastly, we clustered the 96 residues into six antigenic epitopes, which we named H9_A, H9_B, H9_C, H9_D, H9_E, and H9_F, using the K-means clustering algorithm and calculating Euclidean distances between each residue (as demonstrated in Fig. S3 and Table S2).

A feature-based model for antigenic relationship prediction

To predict the antigenic relationship between two viruses, we have developed a machine-learning approach. The method consists of three steps, which are described below:

We selected 14 features that are likely to be informative for predicting the antigenic relationship between H9N2 viruses.[35–37] There are six antigenic epitope features, five physicochemical properties of amino acids features, the receptor binding site feature, and two kinds of glycosylation site features. We measured the differences in the 14 features between HA pairs. As for HA epitope features, we calculated the number of differences in each epitope for each pair of strains. As for the physicochemical properties of amino acid features, we calculated the average change between strain pairs. We obtained the quantitative values of the five physicochemical properties for the 20 amino acids from the Amino Acid index database[38] (the database entries FASG890101, GRAR740103, ZIMJ680104, CHAM820101, and JANJ780101 recorded quantitative descriptions of hydrophobicity, volume, charge, polarity and ASA of the 20 amino acids, respectively). If the number of residues with amino-acid changes were greater than three, only the top three residues with

maximal changes were considered in the calculation. As for the receptor binding site feature, we used the structure of A/swine/Hong Kong/9/98 as a template (PDB: 1JSD) to calculate the average of the shortest Euclidean distances between the residues with amino-acid changes and the three receptor-binding regions. The Euclidean distance between two residues was calculated between their respective Carbon-α atoms. If more than three mutations occurred, only the top three shortest Euclidean distances were considered in the calculation. As for glycosylation site features, we calculated the number of changed glycosylation sites as predicted using NetNGlyc[39] and NetOGlyc.[40]

To construct the antigenic prediction model, 14 relevant features were used as model features, and the antigenic relationships (similarity or distinction) of HI pairs were treated as model labels. Five classic machine-learning classification models were constructed by the Python package sklearn to predict the relationship between each strain pair, namely Logistic Regression, Support Vector Machine, K-Nearest Neighbor, Random Forest, and XGBoost. To validate the model, 70 % of the collected HI pairs were randomly selected as the training set, and the remaining 30 % were used as the test set, and a five-fold cross-validation was performed.

The accuracy, precision, recall, and F1 score were introduced to evaluate the model performance as formula (2), (3), (4), and (5) illustrated.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$Precise = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1 = 2 \times \frac{Precise \times Recall}{Precise + Recall} \tag{5}$$

Here *TP* indicates true positive samples, *TN* indicates true negative samples, *FP* indicates false positive samples, and *FN* indicates false negative samples.

Antigenic network construction

Based on the classification probabilities determined by the model, we calculated the logarithm of the ratio of the probability of being predicted to be antigenically similar to the probability of being predicted to be antigenically different to denote the extent of antigenic similarity. If the logarithm of the ratio is > 0, the antigenic relationship of the two viruses is considered antigenically similar; otherwise, it is considered antigenically distinct. The greater the logarithm of the ratio, the more likely it is that the two viruses are antigenically similar. Cytoscape[41] was used to construct and visualize the antigenic correlation network. All viral pairs predicted to be antigenic similar were connected to generate the network. The nodes in the network are H9N2 strains, and the edge between the two nodes refers to the antigenic similar relationship. Subsequently, to infer antigenic clusters, viruses with similar antigenicity from the ACnet were grouped using the Markov clustering algorithm (MCL),[42] which is designed for network clustering (Fig. S6).

Identification of the cluster-transition determining sites

In this study, we identified antigenic critical positions by considering positions with both highly antigenic discriminating scores and high genetic diversity.[43] To measure the genetic diversity of an amino acid position i ($i = 1 \sim 317$) with 20 amino acid types, we used Shannon entropy and it is defined as:

$$H(i) = -\sum_{i=1}^{20} P(X_i = T)\log(P(X_i = T)) \tag{6}$$

where $P(X_i = T)$ represents the probability of the position i with amino acid type T. Information gain measures the score of amino acid positions on the HA1 and is used to distinguish between two adjacent antigen clusters. The IG of the position i associated with the antigenic cluster Y is defined as

$$IG(i, Y) = H(Y) - H(Y \mid i) \tag{7}$$

$H(Y)$ is the entropy of antigenic cluster Y (i.e., adjacent cluster C_1 and C_2), $H(Y)$ is defined as

$$H(Y) = -\sum_{T \in \{C_1, C_2\}} P(Y = T)\log(P(Y = T)) \tag{8}$$

$H(Y|i)$ is the conditional entropy of Y when the position i is given, $H(Y|i)$ is defined as

$$H(Y|i) = -\sum_{A \in \{M,N\}} P(Z_i = A)\log(P(Y \mid Z_i = A)) \tag{9}$$

$P(Z_i = A)$ is the probability when the amino acid position i is in state A. $P(Y|Z_i = A)$ is the probability of Y when the amino acid position i is in state A.

For example, in the process of antigen cluster transfer from SS94 to JX13, there are 4073 sequences of antigenic cluster SS94 and 1338 sequences of antigenic cluster JX13. The number of mutations and non-mutations at position 150 are 1890 and 3521. Among the 1890 mutation sequences, the number of cluster SS94 and cluster JX13 are 552 and 1338, respectively; among 3521 non-mutation sequences, all the sequences belonged to antigenic cluster SS94. According to these data, we can calculate that $P(Z_{150} = M)$ is 0.35 and $P(Z_{150} = N)$ is 0.65. Finally, we obtained $H(Y|i) =$ 0.09 and $H(Y)$= 0.81, which make the $IG(150, Y)$= 0.72. The values of information gain and entropy of 317 HA positions are normalized in the range from 0 to 1. Given that single amino acid substitutions at critical positions can lead to antigenic changes,[44] we defined “high entropy” as an entropy value greater than 0.8 and “high information gain” as an information gain value greater than 0.7. These thresholds were established to identify 1–3 amino acid positions that are significant for antigenic variation.

PREDAC-H9-based vaccine strain recommendation strategy

For H9N2 AIVs, we predicted pairwise antigenic relationships for any two viruses using the PREDAC-H9 approach and connected viruses with similar antigenicity to construct the ACnet. Subnetworks containing strains from the same year were extracted and the degree value (D_i) was calculated for each strain in the subnetwork, which represents the number of strains similar to strain i in antigenicity. All D_i values were sorted in descending order to select the recommended strain 1 with the largest antigenic coverage rate. The antigenic coverage rate C_i is defined as follows:

$$C_i = \frac{D_i}{N} \tag{9}$$

where N represents the total number of strains in a given year. The recommended strain 1 will be considered as the candidate vaccine strain. When encountering the recommended strain 1 with a low antigenic coverage rate, it is possible to remove the recommended strain 1, and its antigenically similar strains from the subnetwork, and the abovementioned steps were repeated to choose recommended strain 2 with the maximum coverage rate. These two strains will be considered as the vaccine strain candidate combination for the current year. If the antigenic coverage rate of the combination remains low, the recommended strain 3 can be selected, and so on, in a similar manner. In this study, we have provisionally

established the cut-off threshold for antigenic coverage at 80 %. Instances falling below this threshold are deemed to necessitate further recommendation and evaluation.

Phylogenetic analysis

The phylogenetic tree was constructed using Fasttree[45] version 2.1.11 based on the maximum likelihood model and was visualized using Figtree[46] version v1.4.4.

Results

Model and determinants for antigenic relationship prediction

In order to simulate the antigenic pattern of influenza A (H9N2) virus, we refined the PREDAC method[47] for H9N2 AIVs, denoted as PREDAC-H9, to rapidly identify the antigenic variants of the H9N2 virus. The workflow of the method is illustrated in Fig. S1. Five machine learning models integrating 14 structural and physicochemical features of HA sequences were tested to maximize prediction accuracy. These 14 features could be divided into 4 groups, namely newly inferred epitopes (H9_A, H9_B, H9_C, H9_D, H9_E and H9_F), physicochemical properties (Hydrophobicity, Polarity, Charge, Volume and Accessible surface area (ASA)), glycosylation sites (N-Glycosylation, O-Glycosylation), and receptor binding sites (RBS), which have been demonstrated to correlate with influenza antigenicity in past studies.[47] Model training, verification, and prediction were conducted on the divided training set and test set. In order to compare the performance among the machine learning models, we calculate the prediction accuracy, precision, recall, F1 score, and AUC values of each model (Table 1), and the detailed ROC curves are depicted in Fig. S2. In conclusion, the nonlinear ensemble method XGBoost achieved an accuracy of 81.1 %, precision of 81.4 %, recall of 81.3 %, and F1 score of 81.1 % on the test dataset, and we take this method's results as the final output. The XGBoost algorithm is a scalable decision tree-based boosting algorithm that is an ideal candidate for nonlinear, sparse, and class-imbalanced classification data.[48] It is widely used to solve problems regarding regression, classification, and ranking.

To investigate how the 14 structural and physicochemical characteristics contribute to the antigenic variation of H9N2 AIV, feature importance analysis was performed using Shapley Additive exPlanations (SHAP) value. The epitopes (49.44 %) were considered the main factors contributing to the altered antigenicity of the H9N2 AIV, while other groups, physicochemical properties, glycosylation sites, and RBS, contributed 34.34 %, 10.60 %, and 5.62 %, respectively. Epitope features, specifically H9_C, H9_B, and H9_F, were found to contribute most to the model. This highlights the crucial role of antigenic epitopes in predicting antigenic variation and is consistent with the previous research.[47] Among epitope features, H9_C exhibited the strongest predictive power in antigenic variation. Conversely, epitope H9_A demonstrated the weakest predictive ability. This discrepancy may stem from the fact that H9_C encompasses multiple amino acid positions and is located at the head of the HA protein, while H9_A has fewer amino acid positions and is situated furthest from the head of the HA protein (Fig. S3). Among the physicochemical property features, the polarity of the amino acid was considered the most important (Fig. 1a).

In the prediction of antigenic relationships between H9N2 AIVs, the effective SHAP values (mean SHAP value) of six epitopes features (H9_A, H9_B, H9_C, H9_D, H9_E, and H9_F) and three physicochemical properties of amino acids (volume, charge, and polarity) were found to be less than zero. These features have a significant negative impact on the prediction of antigenic variation, and the larger the values of these features, the greater the likelihood that they will be predicted as variants in the antigenic relationship model. Conversely, hydrophobicity, ASA, glycosylation sites, and RBS features have a significant positive impact, and the larger the values of these features, the greater the likelihood of predicting to be similar (Fig. 1b).

Table 1
Model performance metrics.

Models	Accuracy	Precision	Recall	F1_score	AUC
SVM	0.716	0.731	0.720	0.714	0.790
Logistic Regression	0.739	0.749	0.741	0.737	0.812
KNN	0.782	0.791	0.784	0.781	0.857
Random Forest	0.788	0.799	0.791	0.787	0.859
XGBoost	0.811	0.814	0.813	0.811	0.894

SVM, Support Vector Machine; KNN, K-Nearest Neighbor; AUC, Area Under the Receiver operating characteristic Curve.

Antigenic variation of the avian influenza A (H9N2) virus from 1966 to 2022

In order to better understand the antigenic variation and transmission of H9N2 viruses globally, a total of 10,289 HA sequences of H9N2 viruses isolated from different countries and regions from 1966 to 2022 were downloaded. Using the PREDAC-H9 model, the antigenic relationships between all pairwise combinations of H9N2 viruses were predicted, and the similar antigenicity virus pairs were connected to construct an ACnet (Fig. 2a). Then, the predicted antigenic clusters could be identified from the ACnet. A total of 24 antigenic clusters were identified, among which 10 major antigenic clusters included more than 10 % of the viruses in at least one year. As shown in Fig. 2a and b, the ACnet and phylogenetic tree vividly depict the antigenic patterns of H9N2 viruses globally.

We used the short name of the earliest strain that belongs to every major antigenic cluster as the name of this antigenic group (WC66, SS94, CB97, IR98, DE03, SK06, JS13, JX15, GZ16 and GD17). In terms of the four traditional global epidemic lineages of H9 (h9.1-American lineage, h9.2-Y439 lineage, h9.4-G1 lineage and h9.3-BJ94 lineage), each lineage has at least one corresponding independent antigenic cluster. The WC66 antigenic cluster is the only major antigenic cluster in the h9.1-American lineage, and even though it is the oldest prevalent lineage, no significant antigenic variation has been observed yet, probably due to the low prevalence and nearly only transmission in wild birds.[31] The antigenic clusters DE03 and SK06 both belong to the h9.2-Y439 lineage, but there is no direct evolutionary relationship between them. In the prevalence of the h9.4-G1 lineage, two distinct antigenic clusters can be observed, CB97 and IR98. While in the h9.3-BJ94 lineage, which has the largest number of strains, multiple antigenic drifts occurred. The SS94 antigenic cluster appeared earliest but continued to be prevalent until now. Based on the SS94 cluster, four novel antigenic clusters, namely, JS13, JX15, GZ16, and GD17, gradually evolved and became prevalent.

To investigate the detailed antigenic patterns in different regions, we also depicted the antigenic dynamics of AIV in Asia, Africa, Europe, and the Americas from 1966 to 2022 (Figs. 2c and S4). Notably, significant variations were observed in the antigenic patterns of H9N2 AIV across these continents. Among the four continents, Asia had enough sampled sequences and exhibited the most intricate antigenic pattern, characterized by the presence of all the 10 major antigenic clusters with regional heterogenicity (Fig. 2c). In Africa, the predominant antigenic cluster was IR98 (Fig. S4). Despite the low number of strains, the Americas and Europe experienced a long-term predominance of antigenic cluster WC66 (Fig. S4). Among regions in Asia, China played a pivotal role in driving the antigenic dynamics with 9 major antigenic clusters (Fig. 2c). There was co-circulation of both dominant and non-dominant antigenic clusters in Asia and China. Moreover, we have observed the re-emergence of

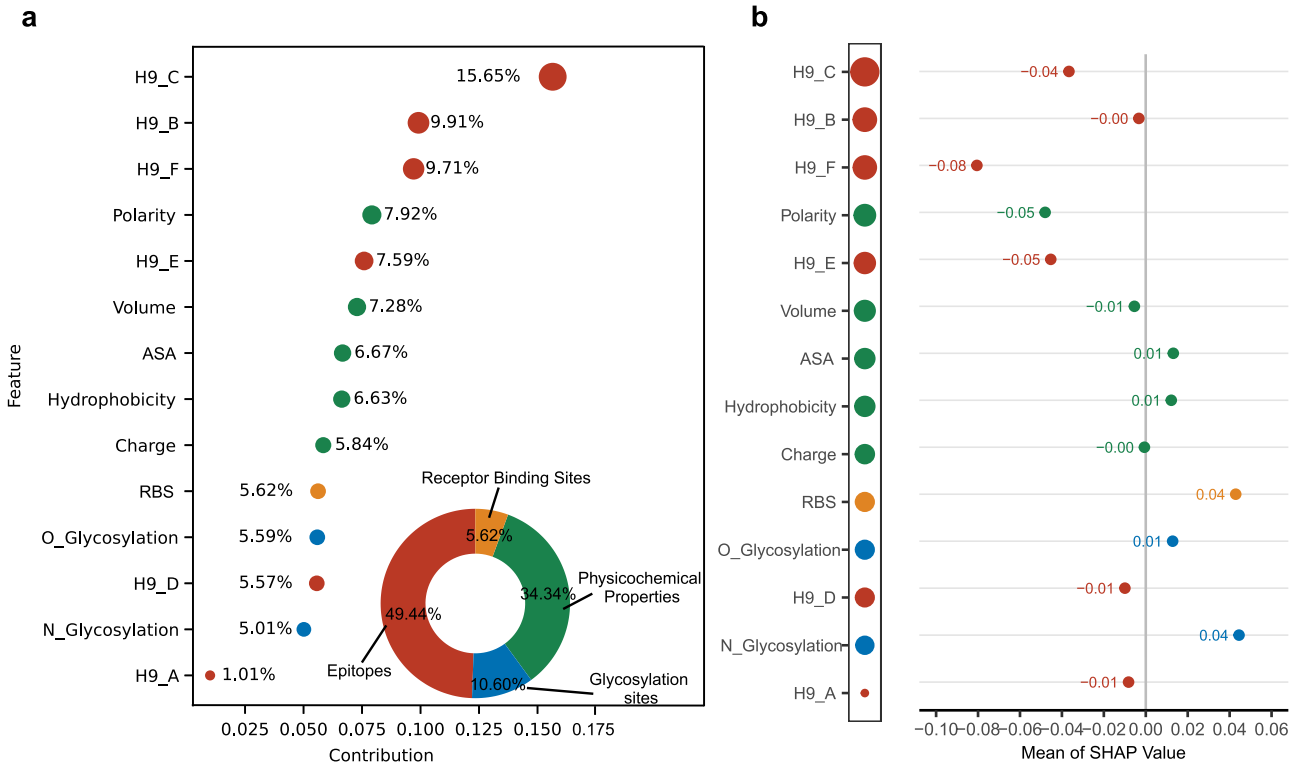

Fig. 1. Feature importance based on SHapley Additive exPlanations (SHAP) values. (a) Ranking of the contributions of 14 features and the importance of the contribution of 4 feature groups. (b) Effective SHAP values (mean of SHAP values) for each feature.

the old dominant antigenic cluster SS94 after being replaced by the new antigenic cluster JS13, regaining its dominance over a period of time (Fig. 2c). South Korea, Vietnam, and Bangladesh with sufficient sequences were also investigated. South Korea is mainly prevalent in antigenic clusters WC66, SK06, and SS94 (Fig. 2c). Meanwhile, Vietnam was similar to China, which showcased four out of the five antigenic clusters within the h9.3-BJ94 lineage, while Bangladesh predominantly experienced the antigenic cluster IR98 (Fig. 2c).

Cluster-transition determining sites

To investigate the amino acid positions that play a role in antigenic cluster transition, the Shannon entropy and information gain (IG) were calculated for each amino acid on HA1 during the inter-cluster transitions. Amino acids with high IG at a specific position imply that the position is highly correlated with antigenic variation. Amino acids with high entropy imply that the position is frequently mutated in the dataset and has high genetic diversity. Given the diversity of antigenicity and the adequacy of data, we have focused our analysis on the determining sites within the antigenic evolution process of the h9.3-BJ94 lineage. Fig. 3 shows the relationship between IG values and entropies of 317 amino acid positions of HA1 during four antigenic transitions within the h9.3-BJ94 lineage. All positions can be divided into four groups based on the values of IG (degree of antigenicity) and entropy (genetic diversity). Positions with high IG and high entropy (i.e. region I) are considered to be critical in the process of antigenic cluster conversion.

In the four antigenic transitions, a total of 6 out of 317 positions were identified as potential determining sites of antigenicity, all located within the six defined antigenic epitopes: two in H9_B (180, 183), two in H9_C (150, 153), one in H9_F (131) and one in H9_A (48). Specifically, three high entropy and high information gain positions, 48, 150, and 180, were identified as potential determinants of transition from SS94 to JS13 (Fig. 3A); a single substitution of D183G led to the transition from JS13 to JX15 (Fig. 3B). Three high entropy and high information gain positions, 150, 153, and 183, were regarded as potential determinants of transition from JS13 to GZ16 (Fig. 3C); substitutions at positions 131, 150, and 183 resulted in the transition from JS13 to GD17 (Fig. 3D). Positions 48, 131, 153, and 180 were only involved in once cluster transition, while positions 150 and 183 were involved in three times. Furthermore, a single residue substitution could lead to cluster transition. In addition, the phylogenetic analysis verified the cluster-transition crucial positions identified in our results. For example, at position 183, the amino acid at this position was almost purely aspartic acid (D) at cluster JS13, while the amino acid at this position is almost all mutated to asparagine (N) when the cluster transfer to GZ16 (Fig. S5).

Strategy for vaccine strain recommendation

Considering the antigenic diversity of the H9N2 AIV in China and its significant role in the global spread of avian influenza, a vaccine strain recommendation strategy was developed using China as an illustrative example. This strategy aims to identify, on an annual basis, a strain or combinations of strains with a high antigenic coverage rate as potential candidate vaccine strains. Table 2 provides a comprehensive overview of the candidate vaccine strains (designated as "recommend strain 1") offered each year from 2010 to 2022, along with their corresponding antigenic clusters, as well as their antigen coverage rate. Furthermore, for years with strains antigenic diversity, recommended strain 2 and strain 3 are provided, along with the corresponding cumulative antigenic coverage rate. Unfortunately, the H9N2 AIV sequencing data has experienced a significant decline from 2019 to 2022, thereby resulting in suboptimal performance of the vaccine strain recommendation strategy. Nonetheless, the vaccine strain recommendation was quite effective during the period from 2010 to 2018. Specifically, the recommended vaccine strains for the years 2010 and 2011 exhibited an impressive antigenic coverage exceeding 95 % against the detected viral strains for each respective year.

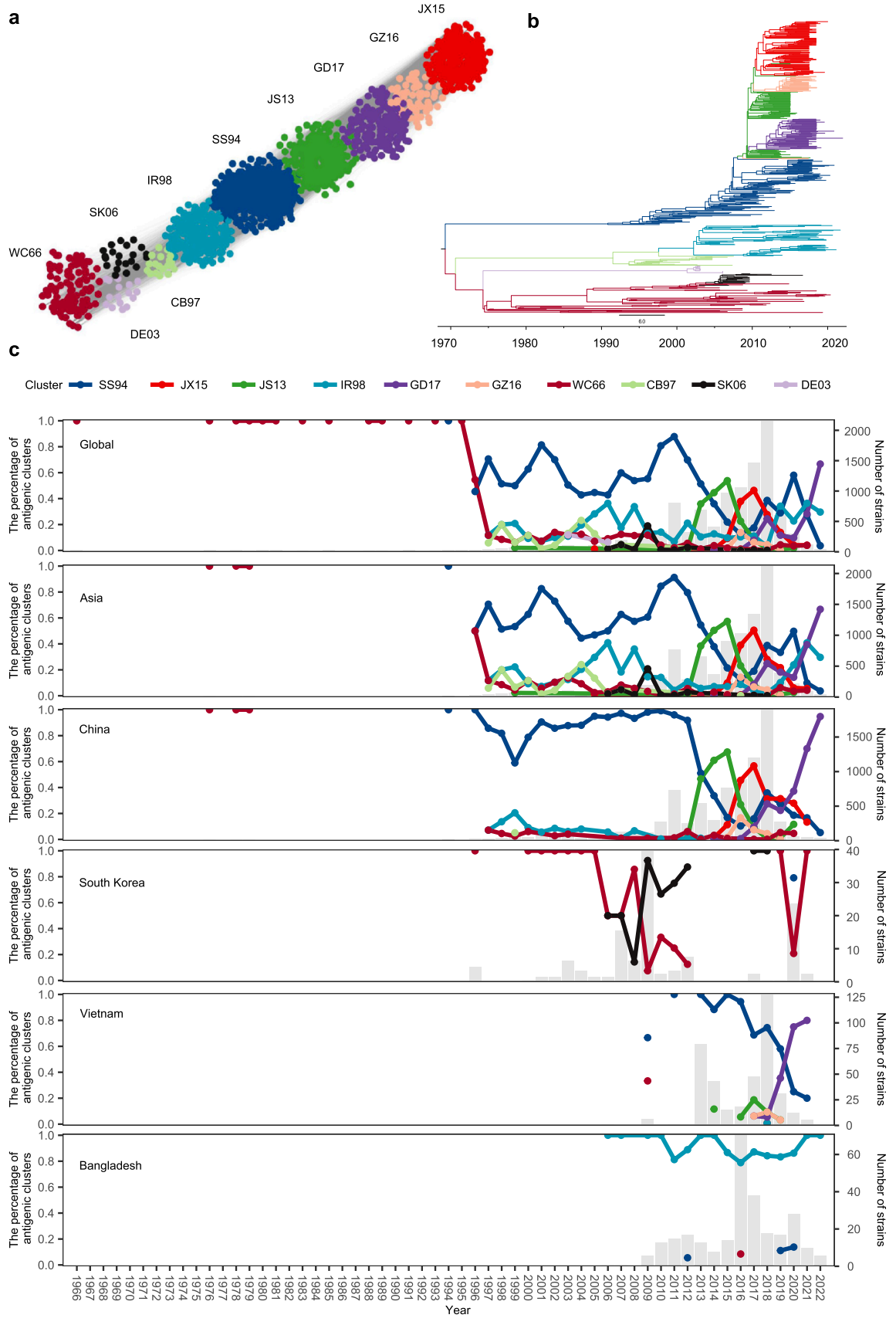

Fig. 2. Global antigenic evolution of the H9N2 virus. (a) Predicted ACnet and antigenic clusters for the sampled viruses. The predicted antigenic clusters are colored and named according to the representative strains contained in the clusters. (b) Phylogenetic tree of HA protein sequences of H9N2 AIVs. Different color represents a different antigenic cluster. (c) Comparison of antigenic patterns of H9N2 viruses from 1966 to 2022. Dynamic changes in the percentage of antigenic clusters were recorded by year for global, Asia, China, South Korea, Vietnam, and Bangladesh. Different colors represent different clusters. The number of sequences were shown in gray bars (JX, Jiangxi; JS, Jiangsu; IR, Iran; GD, Guangdong; GZ, Guizhou; WC, Wisconsin; CB, Chiba; SK, South Korea; DE, Delaware).

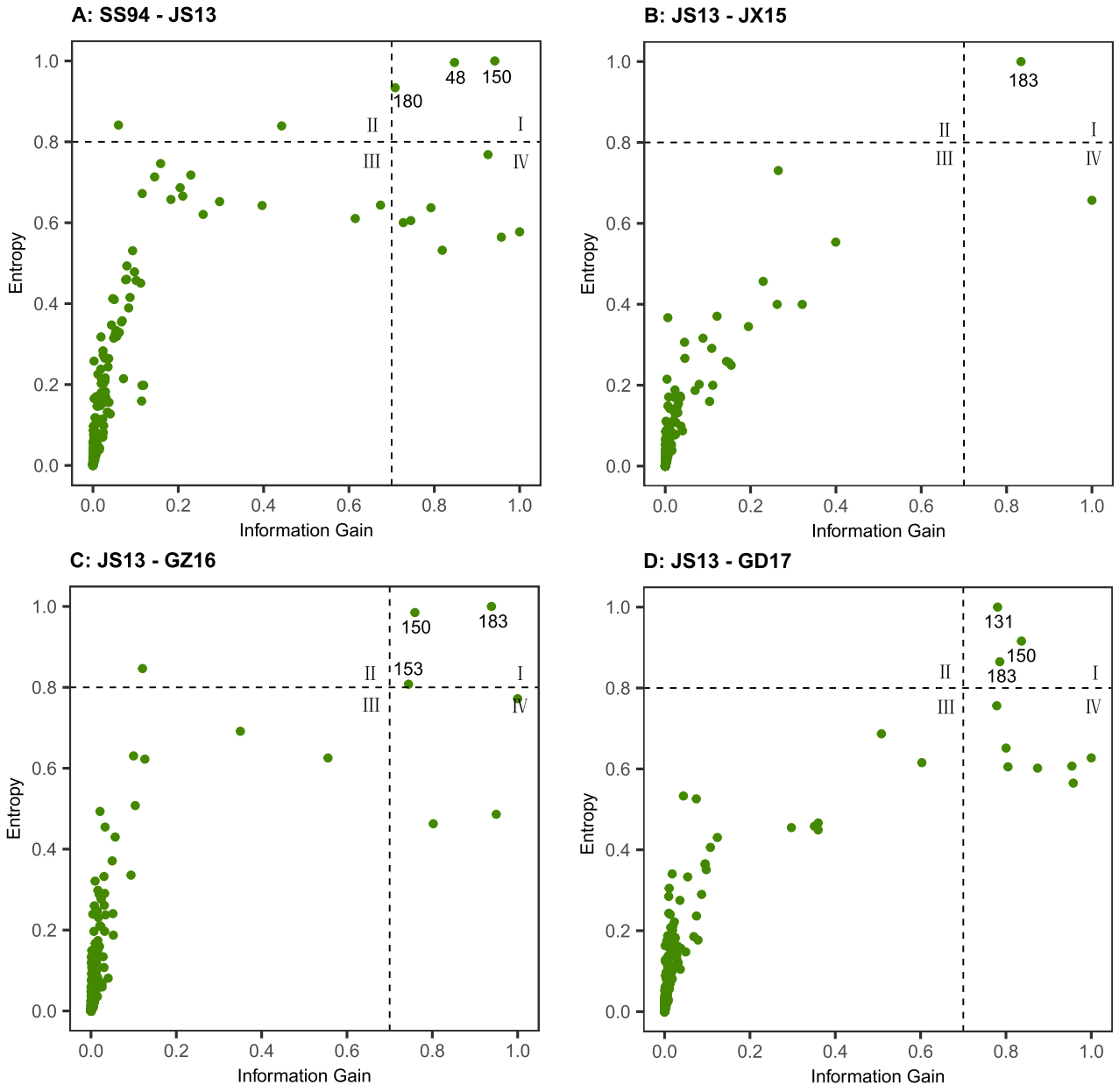

Fig. 3. The entropy and information gain of 317 amino acids on HA1 protein. (A) Entropy and information gain information of 317 amino acid sites in cluster SS94 to cluster JS13. "High entropy" was defined as an entropy value greater than 0.8 and "high information gain" was defined as an information gain value greater than 0.7. Points marked in the first quadrant are sites with high entropy and high information gain. (B) Entropy and information gain of transition of 317 amino acids on HA1 protein in cluster JS13 to cluster JX15. (C) Entropy and information gain of transition of 317 amino acids on HA1 protein antigenic cluster in cluster JS13 to cluster GZ16. (D) Entropy and information gain of transition of 317 amino acids on HA1 protein antigenic cluster in cluster JS13 to cluster GD17.

Since 1998, China has implemented H9N2 AIV commercial vaccine immunization in poultry to mitigate economic losses. We conducted calculations on the antigenic coverage rate of commercial vaccines. The specific data can be found in Table S1. The results indicate that the antigenic coverage rate of the recommended strains is significantly higher than that of the commercially available vaccine strains during the same period. Vaccines with high antigenic coverage are capable of recognizing and eliciting a more extensive antibody response, thus providing stronger immune protection. By monitoring and evaluating antigenic evolution, we can timely understand the variation of epidemic viral strains and update the vaccine strains annually. This practice helps ensure the sustained effectiveness of the vaccines in preventing viral infections and transmissions, thereby reducing the risk of disease outbreaks and safeguarding human and animal health.

Discussion

It is crucial to understand the antigenic evolution pattern of the H9N2 AIV and to identify the newly emerging antigenic clusters for effective vaccination strategies against avian influenza. Current studies have provided evidence supporting the ongoing antigenic evolution of the H9N2 AIV over time, resulting in the emergence of variant antigenic clusters within the virus.[49–51] However, most research on H9N2 antigenic evolution was based on hemagglutination inhibition assay results, which are subject to limitations in terms of temporal scale, spatial scale and sample size.[50,52–54] In this study, we combine PREDAC-H9 with large-scale HA sequencing data to reveal a comprehensive landscape of H9N2 antigenic variation and distribution (Fig. 2).

Table 2
Potential vaccine strains and their antigenic coverage rate for H9N2 viruses from 2010 to 2022.

Year	Recommended strain 1	S1 cluster	ACR	Recommended strain 2	S2 cluster	Cumu-ACR	Recommended strain 3	S3 cluster	Cumu-ACR
2010	A/Chicken/Shanghai/a/2010	SS94	0.992	/	/	/		/	/
2011	A/Chicken/Shandong/bd1/2011	SS94	0.953	/	/	/		/	
2012	A/Chicken/China/tl01/2012	SS94	0.871	/	/	/		/	/
2013	A/Chicken/Shandong/hk1/2013	SS94	0.709	A/Chicken/Guangdong/gch11/2013	JS13	0.911		/	/
2014	A/Chicken/Guangdong/sic31/2014	JS13	0.613	A/Chicken/Wuxi/sc4315/2014	SS94	0.933		/	/
2015	A/Chicken/Qinghai/q1012/2015	JS13	0.867	/	/	/		/	
2016	A/Chicken/Changzhou/0304/2016	JX15	0.713	A/Chicken/Guangdong/10.21_szbj004-o/2016	GZ16	0.871		/	/
2017	A/Chicken/Jiangsu/j2330/2017	JX15	0.769	A/Chicken/Ningxia/nx2128/2017	JX15	0.858		/	/
2018	A/Chicken/Jiangxi/x2372/2018	GZ16	0.841	/	/	/		/	
2019	A/Chicken/Guangdong/131/2019	GD17	0.515	A/Chicken/Qingyuan/g220/2019	SS94	0.773	A/Quail/Guangdong/257/2019	/	0.879
2020	A/Guizhou/13392/2020	GD17	0.302	A/Mink/China/chick_embryo/2020	JX15	0.558	A/Chicken/China/xiangtan/2020	SS94	0.721
2021	A/Hunan/00468/2021	GD17	0.533	A/Chicken/Shandong/102/2021	SS94	0.667	A/Guangdong/02591/2021	JX15	0.767
2022	A/Chicken/China/ningxia6/2022	GD17	0.737	A/Chicken/Guangdong/f99/2022	GD17	0.842	/	/	/

ACR, antigenic coverage rate; Cumu-ACR, cumulative antigenic coverage rate.

Consistent with the previous research on the prediction of influenza virus antigenicity,[47,55,56] we utilized features such as antigenic epitopes, physicochemical properties of the amino acids, receptor-binding sites, and glycosylation sites to predict antigenic relationships between H9N2 viruses. However, unlike more extensively studied viruses such as H3N2 or H1N1, the antigenic sites of H9N2 viruses have not yet been entirely characterized. In order to improve the accuracy of the antigenic relationship prediction model, we inferred a comprehensive set of antigenic sites based on the HA1 protein structure using the webserver ScanNet,[29,30] and built six antigenic epitopes for H9N2 virus, which played an important role in the model (Table S2 and Fig. 1). We found that the inferred antigenic epitopes H9_C, H9_B and H9_F, located on the head of HA, are in the top three in terms of their contribution to the antigenic variation of H9N2 viruses (Figs. 1 and S3). The importance of epitopes H9_C, H9_B and H9_F in antigenic variation is further supported by the fact that 21 of the 26 residues reported to be involved in antigenic variation were distributed in these three epitopes, significantly higher than the ratio in other epitopes. (Binomial test, P-value = 0.02674).[31–34]

Through an analysis of H9N2 HA sequence data isolated over the years 1966 to 2022, we have identified 10 major antigenic clusters (SS94, JX15, etc.). Based on this finding, we mapped the global antigenic patterns, as well as the patterns specific to different continents and important regions. Unlike human seasonal influenza viruses, H9N2 AIVs have heterogenic regional epidemiological characteristics in poultry, and indeed different antigenic landscapes were observed in most countries. All 5 antigenic clusters under the h9.3-BJ94 lineage arose in China. Consistent with reported antigenic drift events,[17,19,57] antigenic drift cluster, JS13, that occurred around 2013 was accurately identified in our study, partly due to the stricter criteria in the model for determining antigenicity similarity. Multiple clusters were simultaneously cocirculating after 2015, which is supported by other reports.[20,49] Interestingly, the h9.3-BJ94 lineage had developed five major antigenic clusters, whereas the G1 lineage, which was also widely prevalent in poultry, had developed only two major antigenic clusters. The quick antigenic drift of h9.3-BJ94 lineage might be a result of multifactorial effects. It is likely, however, that the long-term vaccine immunity pressure to the h9.3-BJ94 lineage is one of the important reasons.[21]

Going a step further, we identified eight forms of mutations across six amino acid residues during the four serial cluster transitions in h9.3-BJ94 lineage. Consistent with the findings of Yan et al.,[18] our analysis identified mutations in three amino acid residues (A150/D/E, H48Q, and V/A180T) on HA protein. Additionally, we identified five previously unknown forms of mutations, including D150N, I153T, D183N, T131N, and D183G (Fig. 3). These variants, except H48Q, are all located on epitopes B, C, and F that we identified. This result is similar to what has been reported for other influenza virus subtypes, i.e., substitutions near the hemagglutinin receptor-binding site determine antigenic cluster transitions.[44,58,59]

China implemented poultry vaccination to prevent the spread of H9N2 in 1998.[21] However, due to the continuous antigenic variation of the H9N2 viruses and the slow renewal of vaccine strains, mismatches between vaccines and prevalent strains occasionally occurred.[60] Our previous studies have shown that antigenic drift of H9N2 is an important reason for the increase of H9N2 outbreaks in chickens.[16,19] Therefore, H9N2 AIV requires an effective vaccine strain recommendation to rapidly curb its spread and minimize its negative impacts when novel antigenicity strains are prevalent. Our study provides an effective solution for vaccine strain matching and maximizing vaccine coverage. Based on continuous monitoring data, we have identified representative strains annually that exhibit strong antigenic similarity to the circulating viruses, suggesting their broad protective potential as ideal vaccine candidates. It is important to note that the success of our vaccine strain recommendation

strategy is greatly influenced by the quality and timeliness of the monitoring data.

Although our study has made significant discoveries, there are several limitations that should be acknowledged. Firstly, the results are limited by the bias distribution of the sequence data in different regions, and the HI data in different periods and lineages. Unlike seasonal influenza, H9N2 AIVs exhibit strong regional epidemics and highly divergent lineages. High-quality clinical surveillance data is a prerequisite for a well-run model, although current data are enough for the model. Secondly, owing to the inherent uncertainties arising from potential inaccuracies in the experimental data and the inherent predictive accuracy of the model, it is possible that certain viral strains may be misclassified. Despite these issues, the overall classification of antigenic clusters was relatively reliable. The percentage of antigenically similar strains in the HI assay that were successfully clustered is 78.3 %. Lastly, the critical amino acid residues we identified through information entropy and information gain calculations still require experimental validation. Future studies may focus on testing these findings experimentally through functional assays and structural analyses, which could provide additional insights into the mechanisms underlying antigenic variation.

In summary, our study developed a specific model for the rapid identification of H9N2 antigenic variants, along with the elucidation of a comprehensive landscape of H9N2 antigenic patterns. We identified important sites for antigenic cluster transitions, as well as providing better vaccine strain recommendations for H9N2 AIVs, which is of significant importance in the better control and prevention of H9N2 AIVs epidemics.

Funding

This work was supported by the National Key R&D Program of China under Grant 2022YFC2303800, the National Natural Science Foundation of China under Grant 81961128002, the Hainan Science Foundation under Grant 323CXTD377, the National Key R&D Program of China under Grant 2021YFC2301300, the Guangdong Frontier and Key Tech Innovation Program under Grant 2022B1111020006, the Major Program of Guangzhou National Laboratory under Grant GZNL2024A01002, and the Shenzhen Science and Technology Program under Grant ZDSYS20230626091203007.

Author contributions

XD, TJ and JP designed the study. KZ and JD collected the data. KZ performed the analysis. KZ, JD, JZ, PC, XW, YC, ZQ, YZ, XD, TJ and JP interpreted the data. KZ and JD prepared the manuscript. KZ, JD, JZ, PC, XW, WH, YZ, XD, TJ and JP edited the paper. All authors reviewed and approved the submitted manuscript.

Availability of data and materials

The data that support the findings of this study are available in the Supplementary Data.

Declaration of Competing Interest

The authors declare that they have no competing interests. There is no financial/personal interest or belief that could affect their objectivity.

Acknowledgments

We gratefully acknowledge all the authors from the origin laboratories who submitted and shared data, on which this study is based. Supported by BrightWing High-performance Computing Platform, School of Public Health (Shenzhen), and High-performance Computing Public Platform (Shenzhen Campus), Sun Yat-sen University.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jinf.2024.106199.

References

1. Homme PJ, Easterday BC. *Avian influenza virus infections. I. Characteristics of influenza A-turkey-Wisconsin-1966 virus. Avian Dis* 1970;**14**:66–74.
2. Carnaccini S, Perez DR. *H9 influenza viruses: An emerging challenge. Cold Spring Harb Perspect Med* 2020;**10**:a038588.
3. Guan Y, Shortridge KF, Krauss S, Webster RG. *Molecular characterization of H9N2 influenza viruses: Were they the donors of the "internal" genes of H5N1 viruses in Hong Kong? Proc Natl Acad Sci USA* 1999;**96**:9363–7.
4. Nagy A, Mettenleiter TC, Abdelwhab EM. *A brief summary of the epidemiology and genetic relatedness of avian influenza H9N2 virus in birds and mammals in the Middle East and North Africa. Epidemiol Infect* 2017;**145**:3320–33.
5. Bi YH, Li J, Li SQ, Fu GH, Jin T, Zhang C, et al. *Dominant subtype switch in avian influenza viruses during 2016-2019 in China. Nat Commun* 2020;**11**:5909.
6. Adlhoch C, Fusaro A, Gonzales JL, Kuiken T, Mirinaviciute G, Niqueux É, et al. *Avian influenza overview June-September 2023. EFSA J* 2023;**21**:e08328.
7. Yu Q, Liu LQ, Pu J, Zhao JY, Sun YP, Shen GN, et al. *Risk perceptions for avian influenza virus infection among poultry workers, China. Emerg Infect Dis* 2013;**19**:313–6.
8. Huang R, Wang AR, Liu ZH, Liang W, Li XX, Tang YJ, et al. *Seroprevalence of avian influenza H9N2 among poultry workers in Shandong Province, China. Eur J Clin Microbiol Infect Dis* 2013;**32**:1347–51.
9. Yang L, Zhu W, Li X, Bo H, Zhang Y, Zou S, et al. *Genesis and dissemination of highly pathogenic H5N6 avian influenza viruses. J Virol* 2017;**91**:e02199-16.
10. Gao R, Cao B, Hu Y, Feng Z, Wang D, Hu W, et al. *Human infection with a novel avian-origin influenza A (H7N9) virus. N Engl J Med* 2013;**368**:1888–97.
11. Liu D, Shi W, Shi Y, Wang D, Xiao H, Li W, et al. *Origin and diversity of novel avian influenza A H7N9 viruses causing human infection: phylogenetic, structural, and coalescent analyses. Lancet* 2013;**381**:1926–32.
12. Alexander PE, De P, Rave S. *Is H9N2 avian influenza virus a pandemic potential? Can J Infect Dis Med Microbiol* 2009;**20**:e35–6.
13. Malik Peiris JS. *Avian influenza viruses in humans. Rev Sci Et Tech* 2009;**28**:161–73.
14. Webster RG, Govorkova EA. *Continuing challenges in influenza. Ann N Y Acad Sci* 2014;**1323**:115–39.
15. Wang Q, Ju L, Liu P, Zhou J, Lv X, Li L, et al. *Serological and virological surveillance of avian influenza A virus H9N2 subtype in humans and poultry in Shanghai, China, between 2008 and 2010. Zoonoses Public Health* 2015;**62**:131–40.
16. Pu J, Wang SG, Yin YB, Zhang GZ, Carter RA, Wang JL, et al. *Evolution of the H9N2 influenza genotype that facilitated the genesis of the novel H7N9 virus. Proc Natl Acad Sci USA* 2015;**112**:548–53.
17. Zhang N, Quan KJ, Chen ZX, Hu Q, Nie MS, Xu N, et al. *The emergence of new antigen branches of H9N2 avian influenza virus in China due to antigenic drift on hemagglutinin through antibody escape at immunodominant sites. Emerg Microbes Infect* 2023;**12**:2246582.
18. Yan W, Cui H, Engelsma M, Beerens N, van Oers MM, de Jong MCM, et al. *Molecular and antigenic characterization of avian H9N2 viruses in Southern China. Microbiol Spectr* 2022;**10**:e0082221.
19. Pu J, Yin YB, Liu JY, Wang XY, Zhou Y, Wang ZJ, et al. *Reassortment with dominant chicken H9N2 influenza virus contributed to the fifth H7N9 virus human epidemic. J Virol* 2021;**95**:e01578–20.
20. Zhang JH, Huang LH, Liao M, Qi WB. *H9N2 avian influenza viruses: challenges and the way forward. Lancet Microbe* 2022;**4**:E70–1.
21. Dong J, Zhou Y, Pu J, Liu L. *Status and challenges for vaccination against avian H9N2 influenza virus in China. Life* 2022;**12**:1326.
22. Shu YL, McCauley J. *GISAID: Global initiative on sharing all influenza data - from vision to reality. Eurosurveillance* 2017;**22**:2–4.
23. Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. *MAFFT-DASH: Integrated protein sequence and structural alignment. Nucleic Acids Res* 2019;**47**:W5–10.
24. Protein Data Bank. http://www.rcsb.org/pdb/home/home.do.
25. Swiss-Model. https://swissmodel.expasy.org/repository/uniprot/P03457.
26. Neher RA, Bedford T, Daniels RS, Russell CA, Shraiman BI. *Prediction, dynamics, and visualization of antigenic phenotypes of seasonal influenza viruses. Proc Natl Acad Sci USA* 2016;**113**:E1701–9.
27. Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, et al. *Mapping the antigenic and genetic evolution of influenza virus. Science* 2004;**305**:371–6.
28. Igarashi M, Ito K, Yoshida R, Tomabechi D, Kida H, Takada A. *Predicting the antigenic structure of the pandemic (H1N1) 2009 influenza virus hemagglutinin. PloS One* 2010;**5**:e8553.
29. Tubiana J, Schneidman-Duhovny D, Wolfson HJ. *ScanNet: A web server for structure-based prediction of protein binding sites with geometric deep learning. J Mol Biol* 2022;**434**:167758.

30. 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**:730–9.
31. Carnaccini S, Perez DR. *H9 influenza viruses: An emerging challenge. Cold Spring Harb Perspect Med* 2020;**10**:a038588.
32. Kaverin NV, Rudneva IA, Ilyushina NA, Lipatov AS, Krauss S, Webster RG. *Structural differences among hemagglutinins of influenza A virus subtypes are reflected in their antigenic architecture: Analysis of H9 escape mutants. J Virol* 2004;**78**:240–9.
33. Okamatsu M, Sakoda Y, Kishida N, Isoda N, Kida H. *Antigenic structure of the haemagglutinin of H9N2 influenza viruses. Arch Virol* 2008;**153**:2189–95.
34. Peacock TP, Harvey WT, Sadeyen JR, Reeve R, Iqbal M. *The molecular basis of antigenic variation among A (H9N2) avian influenza viruses. Emerg Microbes Infect* 2018;**7**:176.
35. Knossow M, Skehel JJ. *Variation and infectivity neutralization in influenza. Immunology* 2006;**119**:1–7.
36. Wang CC, Chen JR, Tseng YC, Hsu CH, Hung YF, Chen SW, et al. *Glycans on influenza hemagglutinin affect receptor binding and immune response. Proc Natl Acad Sci USA* 2009;**106**:18137–42.
37. Hensley SE, Das SR, Bailey AL, Schmidt LM, Hickman HD, Jayaraman A, et al. *Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift. Science* 2009;**326**:734–6.
38. Kawashima S, Pokarowski P, Pokarowska M, Kolinski A, Katayama T, Kanehisa M. *AAindex: amino acid index database, progress report 2008. Nucleic Acids Res* 2008;**36**:D202–5.
39. Gupta R, Brunak S. *Prediction of glycosylation across the human proteome and the correlation to protein function. Pac Symp Biocomput* 2002;**7**:310–22.
40. Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KT, et al. *Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. Embo J* 2013;**32**:1478–88.
41. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. *Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res* 2003;**13**:2498–504.
42. Enright AJ, Van Dongen S, Ouzounis CA. *An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res* 2002;**30**:1575–84.
43. Huang JW, King CC, Yang JM. *Co-evolution positions and rules for antigenic variants of human influenza A/H3N2 viruses. BMC Bioinforma* 2009;**10**(Suppl 1):S41.
44. Koel BF, Burke DF, Bestebroer TM, van der Vliet S, Zondag GCM, Vervaet G, et al. *Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution. Science* 2013;**342**:976–9.
45. Price MN, Dehal PS, Arkin AP. *FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One* 2010;**5**:e9490.
46. Figtree. http://tree.bio.ed.ac.uk/software/figtree/.
47. Du X, Dong L, Lan Y, Peng Y, Wu A, Zhang Y, et al. *Mapping of H3N2 influenza antigenic evolution in China reveals a strategy for vaccine strain recommendation. Nat Commun* 2012;**3**:709.
48. Chen TQ, Guestrin C. XGBoost: A Scalable Tree Boosting System. *Kdd'16: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*; 2016:785–794.
49. Wang Z, Li HK, Li YH, Wu ZL, Ai H, Zhang M, et al. *Mixed selling of different poultry species facilitates emergence of public-health-threating avian influenza viruses. Emerg Microbes Infect* 2023;**12**:2214255.
50. Wei Y, Xu G, Zhang G, Wen C, Anwar F, Wang S, et al. *Antigenic evolution of H9N2 chicken influenza viruses isolated in China during 2009-2013 and selection of a candidate vaccine strain with broad cross-reactivity. Vet Microbiol* 2016;**182**:1–7.
51. Fusade-Boyer M, Djegui F, Batawui K, Byuragaba DK, Jones JC, Wabwire-Mangeni F, et al. *Antigenic and molecular characterization of low pathogenic avian influenza A(H9N2) viruses in sub-Saharan Africa from 2017 through 2019. Emerg Microbes Infect* 2021;**10**:753–61.
52. Liu YF, Lai HZ, Li L, Liu YP, Zhang WY, Gao R, et al. *Endemic variation of H9N2 avian influenza virus in China. Avian Dis* 2016;**60**:817–25.
53. Zhang Y, Yin Y, Bi Y, Wang S, Xu S, Wang J, et al. *Molecular and antigenic characterization of H9N2 avian influenza virus isolates from chicken flocks between 1998 and 2007 in China. Vet Microbiol* 2012;**156**:285–93.
54. Xia J, Cui JQ, He X, Liu YY, Yao KC, Cao SJ, et al. *Genetic and antigenic evolution of H9N2 subtype avian influenza virus in domestic chickens in southwestern China, 2013-2016. PLoS One* 2017;**12**:e0171564.
55. Liu M, Zhao X, Hua S, Du X, Peng Y, Li X, et al. *Antigenic patterns and evolution of the human influenza A (H1N1) virus. Sci Rep* 2015;**5**:14171.
56. Yao Y, Li X, Liao B, Huang L, He P, Wang F, et al. *Predicting influenza antigenicity from hemagglutintin sequence data based on a joint random forest method. Sci Rep* 2017;**7**:1545.
57. Wang XY, Liu KT, Guo YQ, Pei YR, Chen X, Lu XL, et al. *Emergence of a new designated clade 16 with significant antigenic drift in hemagglutinin gene of H9N2 subtype avian influenza virus in eastern China. Emerg Microbes Infect* 2023;**12**:2249558.
58. Lewis NS, Anderson TK, Kitikoon P, Skepner E, Burke DF, Vincent AL. *Substitutions near the hemagglutinin receptor-binding site determine the antigenic evolution of influenza A H3N2 viruses in US Swine. J Virol* 2014;**88**:4752–63.
59. Rosu ME, Lexmond P, Bestebroer TM, Hauser BM, Smith DJ, Herfst S, et al. *Substitutions near the HA receptor binding site explain the origin and major antigenic change of the B/Victoria and B/Yamagata lineages. Proc Natl Acad Sci USA* 2022;**119**:e2211616119.
60. Park KJ, Kwon HI, Song MS, Pascua PN, Baek YH, Lee JH, et al. *Rapid evolution of low-pathogenic H9N2 avian influenza viruses following poultry vaccination programmes. J Gen Virol* 2011;**92**:36–50.